Structural and Biophysical Characterisation of Sgt1, An Essential Component in the Assembly of the Yeast Kinetochore

STRUCTURAL AND BIOPHYSICAL CHARACTERISATION OF SGT1, AN ESSENTIAL COMPONENT IN THE ASSEMBLY OF THE YEAST KINETOCHORE

Thesis submitted to the **University of London** for the degree of **Doctor of Philosophy**, 2012/2013

~

Oliver Willhöft
School of Crystallography
Birkbeck College, London

The 4.3 Å dataset of the Sgt1 TPR domain was collected by Cara Vaughan (Birkbeck College) at the Diamond Synchrotron Light Source (Didcot, UK) on beamline IO4-1. All mass spectrometry data was collected by Richard Kerr (University College London). I, Oliver Willhöft, hereby declare that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Oliver Willhöft
September 2012

"Men fear thought as they fear nothing else on earth – more than ruin – more even than death.... Thought is subversive and revolutionary, destructive and terrible, thought is merciless to privilege, established institutions, and comfortable habit. Thought looks into the pit of hell and is not afraid. Thought is great and swift and free, the light of the world, and the chief glory of man."

Bertrand Russell

ABSTRACT

Sgt1, a co-chaperone of Hsp90, plays a fundamental role in plant immunity, protein degradation and kinetochore assembly in yeast. Sgt1 has a multi-domain architecture, including an N-terminal tetratricopeptide repeat (TPR) domain. Through its modular architecture, Sgt1 is believed to bridge the interaction between two key structural components of the yeast inner kinetochore, Skp1 and Ctf13, and Hsp90. This interaction recruits Hsp90 to Ctf13 for subsequent activation into a complex that is competent to bind centromere DNA. This is an essential first step in the assembly of the yeast kinetochore, since activation of Ctf13 is a prerequisite for the subsequent binding of all other components of the kinetochore. The aim of this project was to gain biophysical and structural insight into Sgt1 and the basis of its interaction with Skp1, utilising X-ray crystallography and biophysical methods.

Sgt1 exhibits a novel mode of TPR dimerisation in the structure of an Sgt1:Skp1 complex and a second potential mode of dimerisation in the structure of the Sgt1 TPR domain alone. One of these interfaces utilises yeast specific insertions and was confirmed to exist in solution by mutagenesis studies targeting the interface. Biophysical studies show that Sgt1 exists predominantly as a dimer in solution, but truncating its C-terminal domains exposes regions that promote trimer formation.

The structure of the Sgt1:Skp1 complex shows that the interface with Skp1 comprises the concave groove of the TPR domain. Studies of the stoichiometry of Sgt1:Skp1 suggest that the functionally relevant species is 2:1, though the structure shows a fundamental 1:1 interaction. Targeting this interface by mutagenesis disrupts complex formation and show also that a monomer of Sgt1 is sufficient to bind Skp1.

Together, this data is used to present an updated model for the role of Hsp90 in kinetochore assembly.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

The work presented in this thesis and my entire PhD training would not have been possible without the patience and attentive guidance of my supervisor, Cara Vaughan. To her, I am eternally grateful for all that I have learned over the past four years. Also, for their insightful spin on analytical ultracentrifugation I extend my gratitude to Tina Daviter and Robert Sarra at Birkbeck College. Without them, this work would not be the meticulous monstrosity that it is. For helping me tackle the intricacies of X-ray crystallography, I thank Ambrose Cole, Tracey Barrett and Claire Naylor, for his help with our SEC-MALS experiments John Wilson at the National Institute for Medical Research, Alan Purvis for his guidance and counsel, and for their weighty contribution to proving that it all made sense, I thank Konstantinos Thalassinos and Richard Kerr at University College London for their endless toils in mass spectrometry. Each of you is a master of your craft, and I will remember your lessons forever.

And in those grim hours, when the Science Gods made their many crafts seem insurmountable, it was my friends who stood by me. I wish to thank Sarah Connery, François-Xavier Chauviac, Alice Clark, Jasmeen Oberoi, Katie O'Dwyer, Anathe Patschull and Christos Savva for their support, Marc Lloyd, Csaba Orban, Maria Shvedunova and Ateesh Sidhu, for our priceless exchange of scientific and life-related rambles, Carl Doyle, for the inspirational music, and Kirstyn and Ronie Twumasi, to whom I owe more than I could ever repay or words could ever express. I am forever in your debt for being my friends when I needed you most.

While my education has taught me many things, I feel that I have learned more about life from my brother Jörg Willhoft, my sister-in-law Laura King and their three amazing children, Gabriel Han, Tristan Tang and Phoebe Yuan, than I ever could from lectures or textbooks. Their boundless energy and relentless positivism is infectious and humbling, but above all, inspiring. You have

supported me in many ways these past four years, and so, in every metaphysical sense of the words: thank you for being you.

In the midst of this doctoral calamity, someone was always there to be my lifeline. Someone who was my sense and sensibility, and someone who would calm the metaphorical maelstrom whenever it came crashing down around me. Someone who made 'home', feel like home. To Natalie Louise Dawson: thank you for being a part of my life. I could not have done this without you and I am thankful for every microsecond that you have invested in me.

But in all my twenty-six years of life there are only two who have been selfless beyond measure and none who have sacrificed more: my mother and father, Udo and Tuti Willhöft. Neither of them ever stopped believing that I could achieve everything beyond the realms of my imagination, and without them, I would not be where I am today. This work is the product of their belief in me. And for this reason, I dedicate this thesis to them: *die besten Eltern auf der ganzen weiten Welt*.

Für Mama und Papa

ABBREVIATIONS

17-AAG	17-N-Allylamino-17-Demethoxygeldanamycin
Å	Ångstrom
ADP	Adenosine Diphosphate
Aha1	Activator of Heat Shock Protein 90 ATPase 1
AMP	Adenosine Monophosphate
AMPPNP	Adenylyl-imidodiphosphate
ASU	Asymmetric Unit
ATP	Adenosine Triphosphate
AU	Arbitrary Unit
AUC	Analytical Ultracentrifugation
bp	Base Pair(s)
BTB/POZ	BR-C, tlk and bab/Pox Virus and Zinc Finger
Bub1	Budding Uninhibited by Benzimidazole 1
CBF3	Centromere Binding Factor 3
CBF1	Centromere Binding Factor 1
CCAN	Constitutive Centromere Associated Network
CCD	Charged-Coupled Device
CCP4	Collaborative Computational Project Number 4
Cdc37	Cell Division Cycle Protein 37
CDE	Centromere Determining Elements
CEN	Centromere
CENP-'X'	Centromere Protein X (A-E)
Cep3	Centromere Protein 3
CHIP	C-terminus of HSP70-interacting Protein
CK2	Casein Kinase 2
Cpr6	Cyclosporin-sensitive Proline Rotamase 6
CS	CHORD-containing Proteins and Sgt1
Ctf13	Chromosome Transmission Fidelity Protein 13
CTPR3	Consensus Three-Repeat Tetratricopeptide Repeat Protein
CTPR3Y3	CTPR3 (Triple Tyrosine Mutant)
Da	Daltons
Dam1	Duo1 And Mps1 Interacting 1
DNA	Deoxyribonucleic Acid
dRI	Differential Refractive Index
DTT	Dithiothreitol
EDTA	Ethylenediamine Trifluoroacetic Acid
EPR	Electron Paramagnetic Resonance
ESI-MS	Electrospray Ionisation Mass Spectrometry
ERBB2	v-erb-b2 Erythroblastic Leukemia Viral Oncogene Homolog 2
F-box	Protein:Protein interaction motif first identified in Cyclin F
FL	Full Length
FRET	Förster Resonance Energy Transfer
g	Gram
GHKL	Gyrase, Hsp90, Histidine Kinase, MutL
Grp94	Glucose Related Protein, 94 Kilodaltons
GST	Glutathionine S-Transferase
GUI	Graphical User Interface

GyrB	DNA Gyrase B
His_6	Six-Histidine
Hop	Hsp70/Hsp90 Organising Protein
Hsc82p	Heat Shock Protein, 82 Kilodaltons, Constitutively Expressed
Hsp70	Heat Shock Protein, 70 Kilodaltons
Hsp82p	Heat Shock Protein, 82 Kilodaltons, Inducible
Hsp90	Heat Shock Protein, 90 Kilodaltons
Hsp90α	Heat Shock Protein, 90 Kilodaltons, α Isoform
Hsp90β	Heat Shock Protein, 90 Kilodaltons, β Isoform
HtpG	High Temperature Protein G
hrs	Hours
IEX	Ion Exchange (Chromatography)
IMAC	Immobilised Metal Ion Affinity Chromatography
IM-MS	Ion Mobility Mass Spectrometry
Ipl1	Increase in Ploidy 1
IpgC	Invasion Plasmid Gene C
ITC	Isothermal Titration Calorimetry
kDa	Kilodaltons
kg	Kilogram
L	Litre
LB	Luria-Bertani (Broth)
LLG	Log-Likelihood Gain
LRR	Leucine Rich Repeat
MAD	Multiple-Wavelength Anomalous Dispersion
min	Minute(s)
µL	Microlitre
µM	Micromolar
mL	Millilitre
mM	Millimolar
MR	Molecular Replacement
MS	Mass Spectrometry
NCS	Non-Crystallographic Symmetry
Ndc10	Nuclear Division Cycle Protein 10
nM	Nanomolar
NMR	Nuclear Magnetic Resonance
PCR	Polymerase Chain Reaction
PD	Pull Down
PDB	Protein Data Bank
Pfam	Protein Families Database
Plk1	Polo-like Kinase 1
PP5	Protein Phosphatase 5
Ppn	Population
Resi/Resn	Residue
RMSD	Root-Mean-Square Deviation
RNA	Ribnucleic Acid
RPM	Revolutions Per Minute
s-value	Sedimentation Coefficient Value
S	Svedberg Unit
SAC	Spindle Assembly Checkpoint
SAD	Single-Wavelength Anomalous Dispersion

Sba1	Increased Sensitivity to Benzoquinone Ansamycins 1
SCF	Skp1-Cullin-F-box
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SE	Sedimentation Equilibrium
SEC	Size Exclusion Chromatography
sec	Second(s)
SEC-MALS	Size Exclusion Chromatography-Coupled Multi-Angle Light Scattering
Se-Met	Selenomethionine
SGS	Sgt1-Specific
Sgt1	Suppressor of G2 Allele of Skp1
Sgt1ΔSGS	Sgt1, Residues 1-279
SII	Strep-II Affinity Tag
Skp1	Suppressor of Kinetochore Protein 1
Sli15	Synthetically Lethal with Ipl1 15
SV	Sedimentation Velocity
TB	Terrific Broth
TLS	Titration-Libration-Screw
TPR	Tetratricopeptide Repeat
TRAP1	Tumour Necrosis Factor Receptor-Associated Protein 1
UniProt	Universal Protein Resource
WD40	40-Residue Tryptophan/Aspartate Repeat

LIST OF FIGURES

Chapter 1: Introduction

Chapter 3: Biophysical Characterisation of Sgt1 and the Sgt1:Skp1 Interaction

Chapter 4: Structural Analysis of an Sgt1:Skp1 Complex and the Sgt1 TPR Domain

Chapter 5: Discussion and Conclusion

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 THE MOLECULAR CHAPERONE HSP90

The Hsp90 (heat shock protein, 90 kilodaltons) molecular chaperone machinery is integral to the function of eukaryotes, providing a mechanism for proteins to fold into their native state, in turn undermining cellular stress caused by aggregation. Homologues of Hsp90 represent of class of ATPases found in all domains of life, including HtpG in *Escherichia coli*, Hsc82p and Hsp82p in the yeast *Saccharomyces cerevisiae*, and four isoforms in humans, Hsp90α, Hsp90β, Trap1 and Grp94. The proteins that are folded within the scheme of Hsp90 function are referred to as 'clients'. Client proteins typically associate with Hsp90 via co-chaperone molecules, which may affect its N-terminal ATPase activity. The nucleotide-bound state of Hsp90 is critical in altering its conformational equilibrium; the conformational space that is sampled as part of nucleotide turnover allows it to associate with different clients and co-chaperones (Southworth and Agard, 2008).

Hsp90 constitutes 1-2% of the total protein content in the cytoplasm (Borkovich *et al*., 1989), and its expression is upregulated in response to cellular stress, increasing to 4-6% as part of the cellular response to heat in *Drosophila* fruit flies (Crevel *et al*., 2001). Successful folding or activation of client proteins potentiates signaling cascades and other downstream cellular functions. For example, Hsp90 is responsible for the maturation of type I steroid hormone receptors located at the cell membrane or in the cytoplasm, and regulates oestrogen, androgen and progesterone-mediated transcriptional changes. Binding of a steroid hormone to its cognate receptor stimulates its dissociation from Hsp90 and subsequent translocation into the nucleus, where a dimerised steroid receptor complex initiates transcription (reviewed in Norman *et al*., 2004; Jackson, 2012). Hsp90 also modulates its own expression by sequestering the transcription factor Hsf1 (heat shock factor 1) together with the molecular chaperone Hsp70. In response to stress-related signaling, this interaction is broken, leading directly to the

upregulation of the inducible isoform of Hsp90 in eukaryotes to maintain cellular proteostasis (Whitesell *et al*., 2009).

Proteostasis refers to the stability of the protein fraction of the cellular milieu, which is balanced by the quality control processes of protein folding and degradation. Hsp90 contributes to the former by preventing the aggregation of unstable protein folding intermediates and providing structural environments conducive to correct folding (reviewed in Hartl *et al*., 2011). In this role, Hsp90 buffers deleterious mutations which would otherwise produce developmental defects as a result of protein misfolding (Rutherford *et al*., 1998). Through this manner of pleiotropic regulation, Hsp90 perpetuates the inheritance of morphogenic proteins in flies.

Hsp90 influences the folding of its client proteins by presenting specialised structural environments as part of its conformational cycle (discussed in greater detail in Section 1.1.2). Through co-chaperone molecules, these environments can either be arrested or destabilised to promote client protein folding. The co-chaperone p23, which induces a conformational arrest in human Hsp90, is proposed to stabilise a conformation that is conducive to the binding and folding of client proteins (McLaughlin *et* al., 2006). For example, for successful propagation, picornavirus requires the Hsp90:p23 chaperone pair for folding of its capsid protein, P1 (Geller *et al*., 2005). In this context, Hsp90 also performs a protective role in preventing the degradation of P1. Specific disruption of this interaction impairs viral replication.

Conversely, Hsp90 also acts as a platform for the degradation of its client proteins. Degradation is thought to occur following several rounds of unsuccessful folding and is transmitted to the proteasomal pathway through E3 ligases such as CHIP (C-terminus of Hsc70-interacting protein). CHIP competes directly with the adaptor protein Hop (Hsp70/Hsp90 organising protein) for binding to Hsp90 *in vitro*. Under normal conditions, both CHIP and Hop balance the rate of both processes, and degradation only occurs preferentially when the folding activities of the molecular chaperones are no longer adequate to maintain

cellular proteostasis (Kundrat *et al.*, 2010). In fibroblast cell cultures, pharmacological inhibition of Hsp90 function by the anti-tumour antibiotic 17-AAG, results in the preferential degradation of the receptor tyrosine kinase ErbB2 by the E3 ligase c-Cbl (Xu *et al.*, 2002). Therefore, rather than being a passive platform for cellular proteostasis, Hsp90 might specifically recognise numerous E3 ligases to actively control protein turnover and, through its chaperones, determine the folding pathway of its clientele.

1.1.1 Insights into the Structure of Hsp90

Hsp90 comprises three domains: the aforementioned N-terminal ATPase domain, a middle domain for client binding and a C-terminal dimerisation domain (Pearl and Prodromou, 2006) (Figure 1.1).

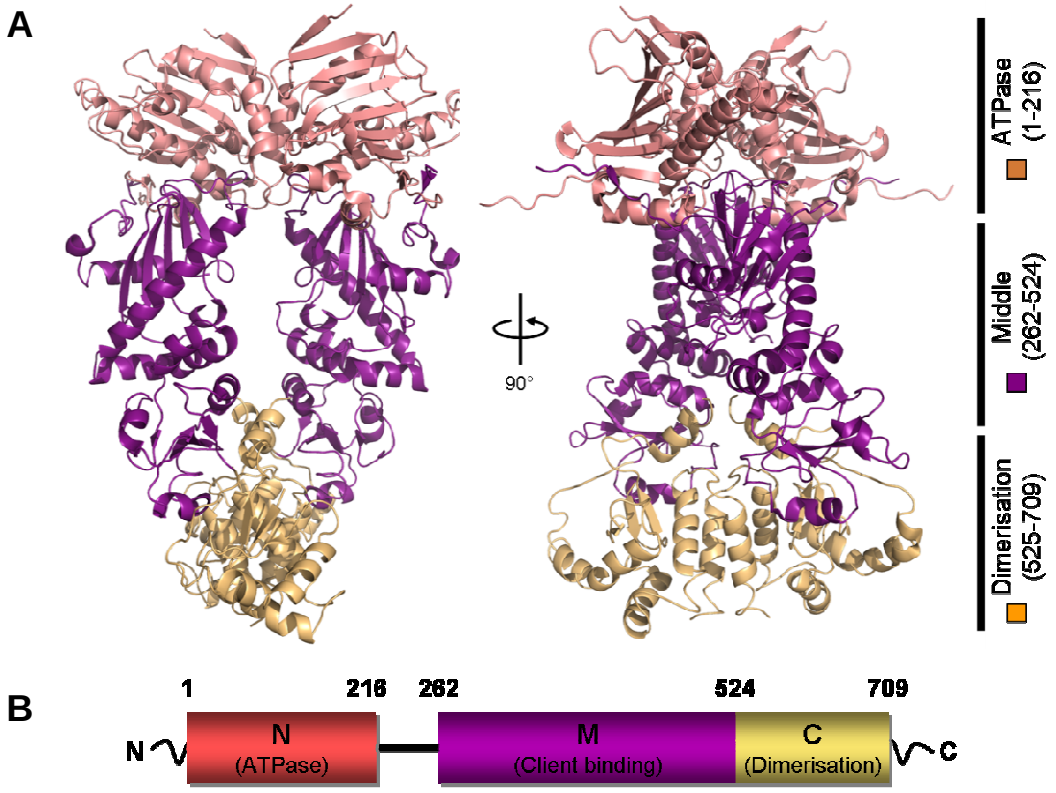


Figure 1.1: **Crystal Structure of Full Length ATP-bound Yeast Hsp90** (PDB ID: 2CG9) – **(A)** The structure of yeast Hsp90 showcases its three domain architecture. The N-terminal ATPase domain (residues 1-216, **Salmon**) binds nucleotide in a manner that influences the self-association state of the N-termini (shown here in the dimerised state). Co-chaperone and client interactions are thought to occur via the middle domain (residues 262-524, **Deep Purple**), which also contributes residues that are crucial to ATP hydrolysis. The C-terminus (residues 525-709, **Light Orange**) contains the site of constitutive dimerisation. **(B)** The linearised domain arrangement of yeast Hsp90.

Several crystallographic studies of *E. coli*, human and yeast Hsp90 homologues have revealed a conserved mode of dimerisation at the C-terminus involving a four-helix bundle formed by helices contributed by adjacent protomers (Figure 1.2) (Ali *et al.*, 2006; Harris *et al.*, 2004). Four residues in yeast Hsp90 (Hsp82) – Leu646, Leu652, Pro661 and Phe664 – are conserved in higher eukaryotes (Pearl and Prodromou, 2006), and form the core of the hydrophobic interaction network that is responsible for the constitutive interaction of adjacent C-termini. This is reflected in the 60 nM K_D for dimerisation (Richter *et al.*, 2001). Truncation of the C-terminus leads to a reduced rate of ATP turnover in human Hsp90, which highlights the importance of dimerisation to the ATP-coupled conformation cycle (Prodromou *et al.*, 2000). However, despite the overall conservation of tertiary structure, the C-terminus is significantly divergent in sequence between species (Pearl and Prodromou, 2006). A Met-Glu-Glu-Val-Asp (MEEVD) motif, which binds a specific class of co-chaperones at the C-terminal extremity of eukaryotic Hsp90 (Young *et al.*, 1998) (discussed later), is notably absent from HtpG, the bacterial homolog. This likely reflects the differences in the co-evolution of Hsp90 and its co-chaperones in the eukaryotic and prokaryotic domains.

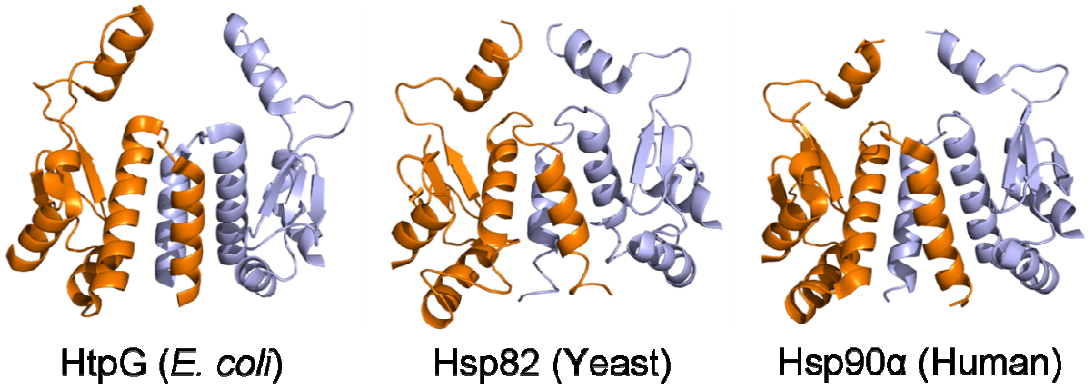


Figure 1.2: **Crystal Structures of the C-terminal Domain of Hsp90 Homologues** (PDB IDs: 1FS8, 2CG9 and 3Q6M) – Constitutive dimerisation with a low nanomolar K_D is brought about by packing of helices at the C-terminus into a four-helix bundle, which involves a hydrophobic core pointing inwards along the axis of the four helices. This mode of dimerisation is conserved in all homologues of Hsp90 and is critical to its conformational cycle.

The structure of the middle segment shows an intra-domain division of two αβα folds conjoined by three helices arranged in a right-handed coil (Meyer *et al.*, 2003) (Figure 1.3). An amphipathic loop structure (residues 328-388 in yeast

Hsp82, 327-340 in human Hsp90α) acts as a critical site for co-chaperone and client protein binding (Meyer *et al.*, 2003), including the kinases Akt (Sato *et al.*, 2000) and Cdk4 (Vaughan *et al.*, 2006), such that mutations in this region inhibit the downstream signaling activity of their respective pathways. Similar to the structural overlap seen at the N-terminus (see later discussion), the middle segments of Hsp90 and other GHKL ATPases highlight an evolutionary divergence in the contributions of this region to their respective catalytic activities.

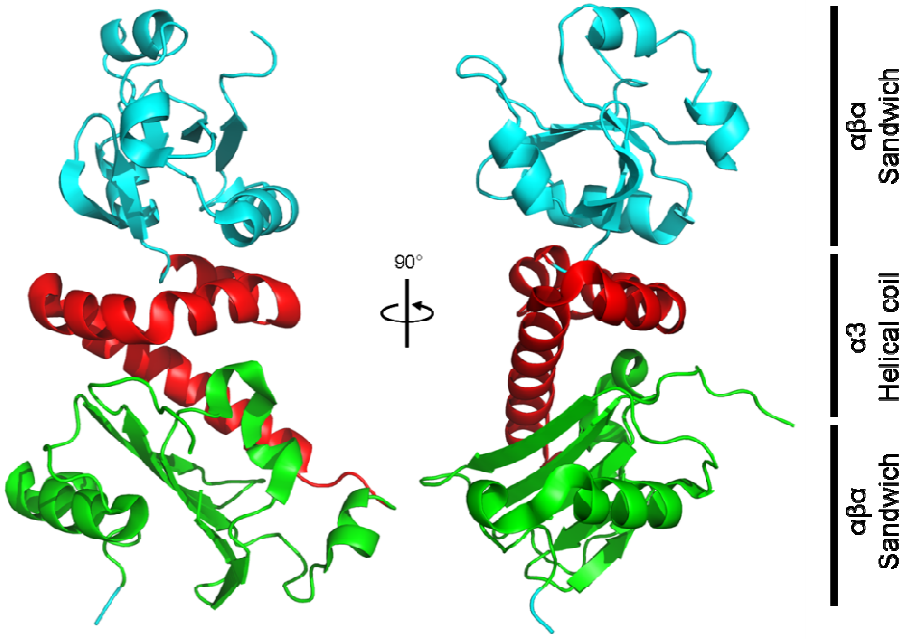


Figure 1.3: **Crystal Structure of the Middle Segment of Yeast Hsp90** (PDB ID: 1HK7) – The middle segment in Hsp90 bridges the gap between the N- and C-terminus, but provides crucial structural features for Hsp90s function, rather than simply being a linker region. The predominantly alpha-helical structure contributes critical residues to the catalytic N-terminal ATPase domain in the conformational cycle of Hsp90. Furthermore, several of its residues are involved in direct interaction with client proteins, similar to the C-terminal domain.

As an example, Aha1, a co-chaperone of Hsp90 that stimulates its ATPase activity, interacts with the middle domain and rearranges residues critical to catalysis by stimulating the projection of a catalytic arginine (Arg380 in yeast) from the middle domain toward the γ-phosphate of ATP bound within the N-terminal active site, which consequently provokes ATP hydrolysis (Figure 1.4) (Meyer *et al.*, 2004). In this manner, Aha1 not only facilitates local changes at the residue level, but also, as will be addressed later, encourages global

conformational changes throughout the chaperone. By analogy, middle segments of MutL and GyrB (two members of the GHKL ATPase family) contribute an equivalent catalytic lysine residue to this process and carry a conserved charged linker region for the binding of DNA clientele, which is akin to the client protein interactions of Hsp90 (Meyer *et al.*, 2003; Pearl *et al.*, 2006). This example highlights the intimate communication between domains of Hsp90, which are critical to understanding its conformational cycle.

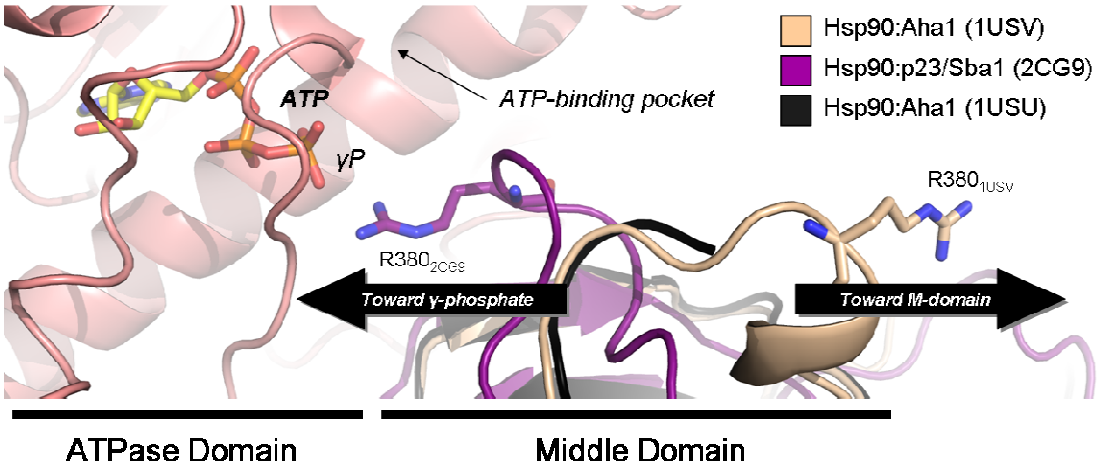


Figure 1.4: **Structural Reshuffling of Hsp90 in the Presence of the ATPase Activator Aha1** (PDB IDs: 1USV, 2CG9 and 1USU) – Aha1 specifically activates the ATPase activity of Hsp90, by reshuffling components that lead to the exposure of residues at the point of nucleotide contact. Specifically, Arg380 rearrangement extends the middle segment (see above) toward the ATP binding site, leading to a faster rate of ATP hydrolysis.

By virtue of its role in propagating cellular cascades, Hsp90 lies at the epicentre of oncogenic transformation and the activation of oncoproteins. As such, one of the key interests in the field is the development of drugs that specifically inhibit Hsp90's function in cancer-related pathways (reviewed in Neckers *et al.*, 2012). The active pursuit of an effective inhibitor has led directly to a wealth of structural information on the N-terminal ATPase domain of Hsp90, which contains the site of nucleotide binding and the hub for conformational changes propagated from the N- to C-terminus. The N-terminus shares a topological overlap with GyrB, specifically in sequences involved in binding and hydrolysis of ATP (Meyer *et al.*, 2003). The first structural characterisation of the N-terminus was in complex with the specific inhibitor geldanamycin (Stebbins *et al.*, 1997), illustrating an αβ sandwich fold consisting of nine alpha-helices and an

eight-stranded anti-parallel beta-sheet that form a nucleotide-binding pocket approximately 15 Å deep (Figure 1.5). Geldanamycin stabilises the N-terminal domain by making van der Waals and extensive hydrogen bond contacts with the beta-sheet base of the hydrophobic pocket, and in turn prevents conformational rearrangements of a segment referred to as the "lid", comprising helices α5 and α6 in all Hsp90 homologues (Stebbins *et al.*, 1997) (Figure 1.5).

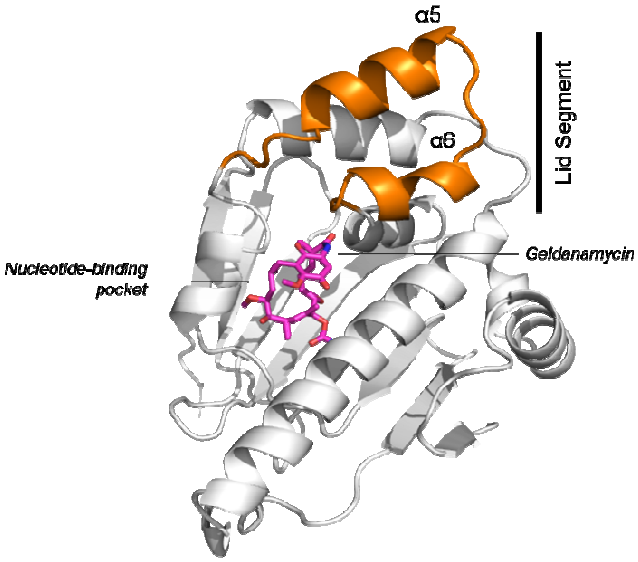


Figure 1.5: **Crystal Structure of the N-terminal Domain Binding Pocket of Human Hsp90 with Geldanamycin** (PDB ID: 1YET) – Geldanamycin makes van der Waal contacts with the beta-sheet base of the ATP binding pocket. Side-chains protruding from the base contribute other stabilising interactions, while the lid segment (**Orange**) is displaced and stabilized at the top of the domain, instead of closing over the active site as occurs during ATP hydrolysis. In contrast to ATP and ADP, geldanamycin does not facilitate a transition from 'open' to 'closed' states. The overall structure of this domain adopts an αβ sandwich fold.

The lid-segment is key to the function of the N-terminal ATPase domain. Deletion of the lid leads to a loss of ATPase activity but not nucleotide binding, suggesting that it is vital for ATP turnover (Richter *et al.*, 2006). This function is facilitated by a movement of the lid-segment over the nucleotide-binding pocket, leading to a "closed" conformation that is stabilised through contacts with ATP (Ali *et al.*, 2006; Vaughan *et al.*, 2009). Mutations that enforce lid closure leads to an increased rate of ATP hydrolysis (Prodromou *et al.*, 2000). Upon ATP hydrolysis, the lid assumes an "open" conformation where helices α5 and α6 point in the opposite orientation to the "closed" conformation and away from the nucleotide-binding pocket (Figure 1.6). Several studies have unveiled different "open" orientations of the lid segment in the Apo and ADP-bound state (e.g. PDB IDs 2IOR/P, 1TBW and 3Q6M) (Immormino *et al.*, 2004; Shiau *et al.*, 2006). While these numerous conformations might indicate an enhanced flexibility in non-ATP-bound states, the conformational space that is sampled is constrained by stabilising bonds emanating from the extreme N-terminus, corresponding to the

first 24 residues in the yeast homolog (Richter *et al.*, 2006). Fluorescence experiments on the conformational dynamics of the lid have shown that its rigidity is an inherent limiting factor in ATP hydrolysis, with lid-closure being the specific rate-limiting step for ATP turnover (Vaughan *et al.*, 2009), though deletion of the 24 N-terminal residues, which increases lid flexibility, results in a loss of ATPase activity (Richter *et al.*, 2006). This data suggests that significant conformational lid flexibility is recalcitrant to ATP turnover and that stabilisation of either the "open" or "closed" structure is important for the controlled catalytic activity of the N-terminal region.

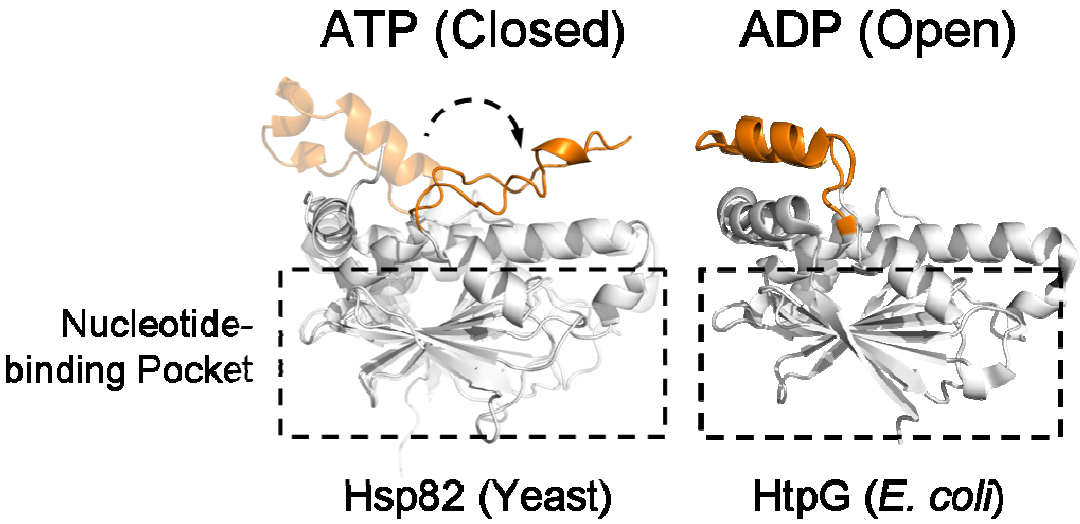


Figure 1.6: **Crystal Structures of the Hsp90 N-terminal Domain with Different Lid Conformations** (PDB IDs: 2CG9 and 2IOR) – Bound nucleotide has the effect of changing the conformation of a lid-segment, which consists of helices α5 and α6 and the adjoining loop. In the ATP bound state, the lid closes over the active site in a manner that exposes hydrophobic residues that were previously buried. This change is partly responsible for N-terminal dimerisation. Upon ATP hydrolysis to ADP, or in the apo state, the lid is considered 'open', pointing away from the N-terminus.

The structural changes associated with lid movement propagate the transient self-association of two adjacent N-termini (Figure 1.7), a function that is critical not only to ATPase activity, but also to the global conformational cycle of Hsp90. Heteromeric Hsp90 dimers comprising one wild type and one monomer lacking a lid-segment, show a drastic increase in ATPase activity and increased self-association of the N-termini (Hessling *et al.*, 2009). Conversely, an Hsp90 hybrid lacking one N-terminal domain shows a reduced rate of ATP turnover (Vaughan *et al.*, 2009). Both observations emphasise the importance of dimerisation to

catalysis and infer that the lid segment plays a role in self-association. Dimerisation buries numerous hydrophobic residues that become exposed upon lid closure and involves a strand-swap of the extreme N-terminus (Figure 1.7). Despite sharing a common mode of dimerisation, Hsp90 homologues show different dimerisation-coupled ATPase activities. Human Hsp90β, for example, shows a 16-fold lower ATPase activity relative to yeast Hsp90 (Hsp82p) (Vaughan *et al.*, 2009).

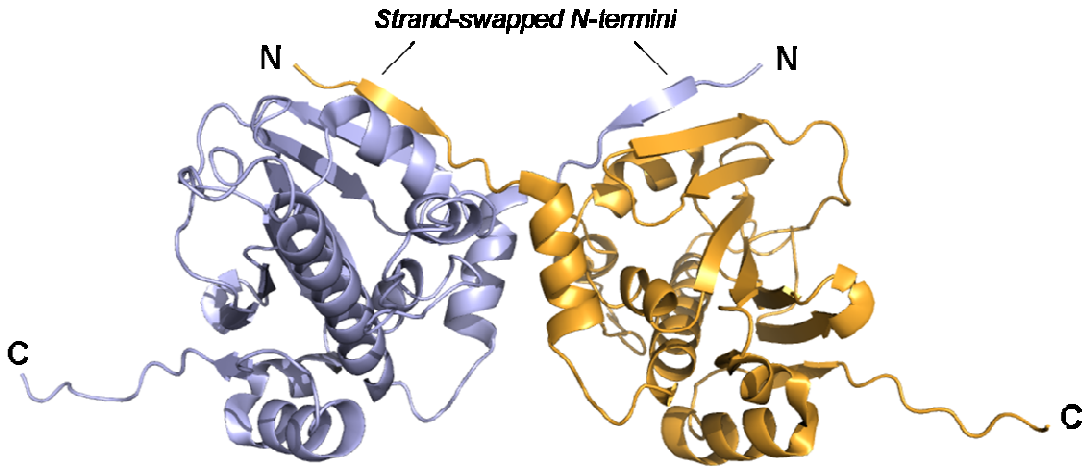


Figure 1.7: **Crystal Structure of the Dimerised N-terminus of Yeast Hsp90** (PDB ID: 2CG9) – Dimerisation of the N-terminal domain is critical to propagating the conformational cycle of Hsp90. Upon ATP binding, newly exposed hydrophobic regions can drive dimerisation and the swapping of strands at the extreme N-terminus. The new hydrophobic core exists at the interface of the two N-termini, though the cleft between the two domains exposes a new hydrophobic surface. This change makes the N-terminally dimerised state energetically unfavourable.

One of the critical aspects of this dimerisation is how such local conformational rearrangements might be signaled from one domain to the next. Clearly, the catalytic function of Hsp90 depends on the integrity of all three domains, from constitutive to transient dimerisation. These changes lead to very distinct transitions in the overall structure of Hsp90, each of which is primed for different client and co-chaperone interactions.

1.1.2 Structural Basis for the Conformational Cycle of Hsp90

The conformational space sampled by Hsp90 is an essential aspect of its function and involves significant domain movements coupled to the nucleotide-bound state of the N-terminal ATPase domain. In a recent assessment, Southworth and

Agard evaluated the conformational cycle of Hsp90 across different species by electron microscopy reconstructions of Hsp90 in different nucleotide-bound states (Southworth and Agard, 2008).

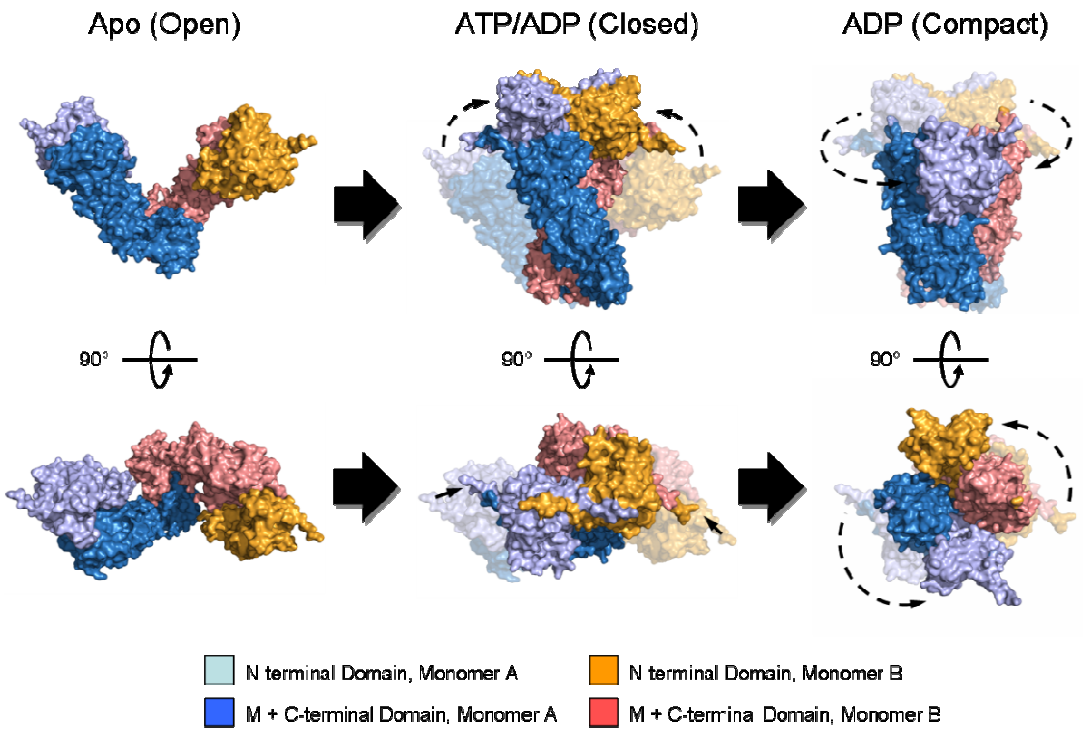


Figure 1.8: **The Conformational Cycle of Hsp90** (PDB IDs: 2IOQ, 2IOP and 2CG9) – The conformational cycle of Hsp90 transitions between three defined states, with the possibility that intermediates also have some functional relevance in the function of the chaperone machinery. In the absence of nucleotide, Hsp90 samples an open or closed conformation, where the N-termini are either adjoined or splayed apart. Upon ATP hydrolysis or ADP binding, the molecule can collapse into a compact state. Each conformational transition entails the exposure of different hydrophobic patches, which carry different affinities for client proteins and co-chaperones.

When bound to Hsp90 from *E. coli* or yeast, the non-hydrolysable ATP analogue AMPPNP brings about a closed but extended conformation of Hsp90, where the N-termini are tightly associated via the previously described interface (Figure 1.8, Closed), while human Hsp90 exhibits little conformational change. In the absence of nucleotide or transiently in the ADP-bound state, all three homologues exist predominantly in the open conformation, where the N-termini are splayed apart (Figure 1.8, Open), but sample the closed state (Southworth and Agard, 2008). A third conformation, in which the N-termini reside on opposite sides of the dimer, is populated in the ADP-bound state, where the re-opened lid segment acts to seal the N-terminal domain against the middle domain (Figure 1.8, Compact) (Shiau *et al.*, 2006). The subtle differences in the

conformational cycle likely represent evolutionary differences in the function of Hsp90 homologues.

One hypothesis for Hsp90 function is based on a mechanism observed for GyrB, where transient N-terminal dimerisation forms a 'molecular clamp' around client DNA, with the hydrolysis of ATP coupled to its release. By analogy, clamping of the N-termini around a client is coupled to its folding and activation (Chadli *et al*., 2000; Prodromou *et al.*, 2000). However, recent single molecule observations have shown that conformational fluctuations do not follow as strict a schematic or are as tightly coupled to ATP hydrolysis as previously believed. Sampling of the open and closed conformations actually occurs on a timescale faster than ATP turnover and global structural fluctuations continue in various nucleotide bound states (Mickler *et al*., 2009). ATP binding to the N-terminal domains also shows significant negative cooperativity, such that binding of ATP to a single site in the open conformation decreases the relative affinity for nucleotide binding to the available site in the second monomer (Ratzke *et al*., 2012). This suggests that both monomers can act independently in the cycle. Such asymmetric stimulation is exemplified by the mechanism of Aha1 which can selectively stimulate the ATPase activity of one monomer (Retzlaff *et al*., 2010). Such selectivity might act as a filter for specifying co-chaperone or client protein binding sites. If a co-chaperone were to have no effect on the ATPase activity (such as the yeast co-chaperone Sgt1 (Bansal *et al.*, 2004; Lee *et al.*, 2004)), then theoretically both halves could potentiate different functions. These observations emphasise the importance of the structural flexibility that underscores the chaperone cycle. In this regard, instead of stimulating a rigid mechanical change, nucleotide binding lowers the energy barrier to particular structural states and increases the conformational space that is sampled in order to enhance client protein and co-chaperone binding. Despite its catalytic activity, the conformational cycle might not be as tightly coupled to ATP-turnover as previously believed.

The purpose of the conformational equilibrium is to rotate through different exposed hydrophobic surfaces for client proteins that require chaperoning, acting more as a 'dynamic clamp' in the scheme of protein folding. In this respect,

those conformations for which structural information is available represent metastable states, that are stabilised by client proteins themselves or by co-chaperones for the purpose of client protein activation. As well as being fundamentally different at the sequence level, differences in the client base of Hsp90 homologues would explain the different conformational preferences seen in the studies of Southworth and Agard. Each transition brings about different binding sites (reviewed in Krukenberg *et al*., 2011). In the apo form, the central cleft contains numerous hydrophobic patches in both the middle domain and at the N-terminus. Binding of ATP stimulates a rotation of helix α1 to expose and consequently bury hydrophobic residues via dimerisation of the N-terminal domains. As a result, the N-termini undergo a small rotation to yield the extended and closed conformation, and produce a region of hydrophobicity at the cleft formed by the dimer interface, while maintaining a site for client protein interactions in the middle domain. Upon lid opening, the changes that occur to bring about the compact conformation bury the hydrophobic regions in a manner that would be conducive to client protein release (Shiau *et al*., 2006). Throughout this cycle, local changes to the energy landscape are communicated across the length of the dimer, with nucleotide, clients and co-chaperones stimulating changes that are thermodynamically favoured in the global scheme of conformations (Morra *et al*., 2009). Several studies have shown that Hsp90 can prevent aggregation and promote folding of native and non-native substrates such as bovine rhodanase and luciferase (Young *et al*., 1997) in a manner dependent on the structural availability of different regions.

In this regard, it is not entirely unreasonable that some regions of the overall structure might host more sites for client protein binding than others, and it is now believed that the middle domain mediates a large portion of these interactions (Hawle *et al*., 2006). It might be expected that some co-chaperones and client proteins will show little to no preference for a specific conformation, while others will require some facilitative action for tight binding to Hsp90. Accordingly, the role of co-chaperones – molecules that dictate client protein binding and contextual function of Hsp90 – is paramount in regulating ATPase activity and conformational rearrangements conducive to activation of clients.

1.1.3 The Role of Co-chaperones in the Hsp90 Chaperone Machinery

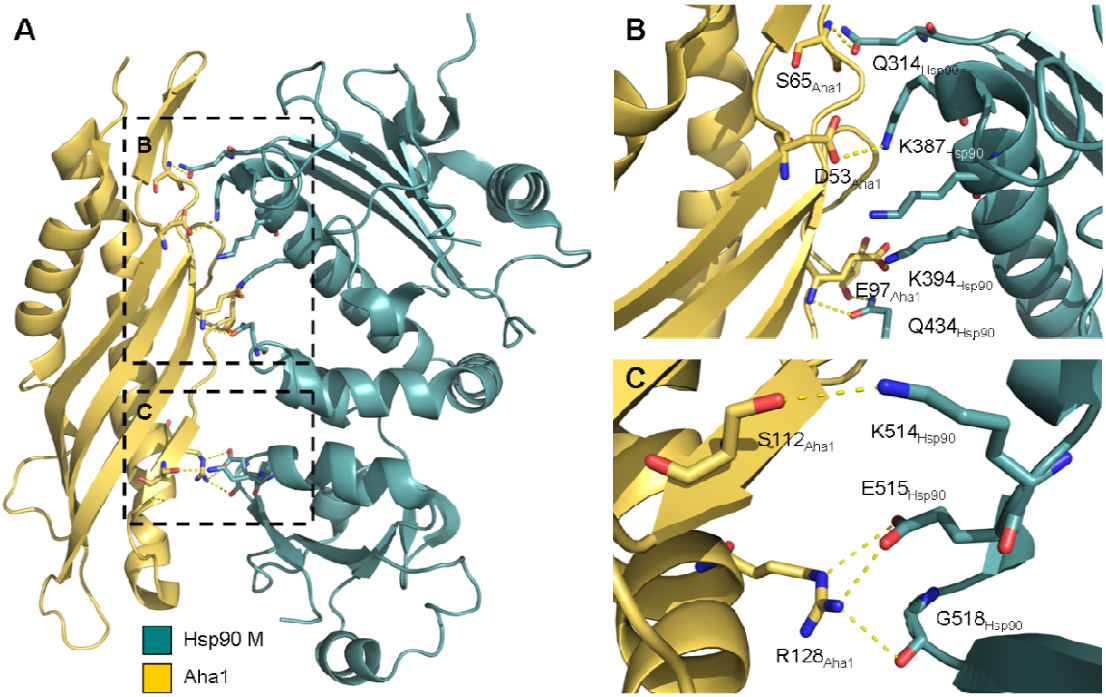


Figure 1.9: **Crystal Structure of Yeast Hsp90 in Complex with Aha1** (PDB ID: 1USV) – **(A)** Crystal structure of yeast Aha1 in complex with a middle-domain construct of Hsp90. Aha1 has an elongated shape that straddles the entire middle domain. This interaction prompts a conformational re-shuffling into the N-terminus of Hsp90, which significantly increases its ATPase activity. **(B)** Close-up of the residues at the interface involving the N-terminal lobe of the middle domain of Hsp90. **(C)** Close-up of the residues at the interface involving the C-terminal lobe of the middle domain of Hsp90.

The joint action of Hsp90 with its co-chaperones can provide a conformational bias towards structural intermediates conducive to client protein binding. The co-chaperone Aha1, for example, activates the ATPase activity of the N-terminal region and increases the Hsp90-mediated activation of clients like tyrosine kinase v-Src *in vivo* (Lotz *et al.*, 2003). However, it has also been proposed that Aha1 modulates the "dwell time" of Hsp90 on its clients and that it destabilises structural intermediates that are conducive to the folding of a disease mutant of CFTR (cystic fibrosis transmembrane conductance regulator) (Kuolov *et al.*, 2010).

Bearing in mind the dynamic structure of Hsp90, the co-chaperone Aha1 effectively elicits an acceleration of the conformational cycle, such that cystic fibrosis mutants of CFTR do not have sufficient time to fold. This mode of

interaction, mediated by the middle domain (Figure 1.9) exemplifies the orchestrated action of chaperone and co-chaperone in the scheme of the molecular chaperone machinery.

On the flipside of upregulated catalytic activity, co-chaperones can also slow the rate of nucleotide turnover to favour other metastable states. Both Hop (Onuoha *et al.*, 2008) and the co-chaperone Cdc37 (Roe *et al*., 2004) are able to arrest the ATPase cycle. Cdc37 binds to the N-terminal domain of Hsp90 and physically prevents closure of the N-terminal lid-segment, while simultaneously projecting an inhibitory arginine residue into the active site to prevent catalysis (Roe *et al*., 2004). Hop, on the other hand, prevents N-terminal dimerisation, a step that is usually coupled to an increase in ATP hydrolysis (Lee *et al*., 2012). By slowing this rate of catalysis, Hop increases the energy barrier to an already unfavourable N-terminally dimerised conformation, and maintains the aforementioned hydrophobic topology conducive to client protein binding.

One feature that is common to a large class of co-chaperones is the tetratricopeptide repeat (TPR) motif (described in greater detail in Section 1.2.1.2). TPR repeats pack together to form a two-faced helical domain and target the C-terminal EEVD peptide motif in Hsp90, with the C-terminal aspartate forming polar interactions with the concave groove of a TPR domain (reviewed in D'Andrea *et al*., 2003) (Figure 1.10). TPR-containing proteins frequently act as adaptors for protein folding, providing a scaffold for the client base of Hsp90, and hold a high affinity for the C-terminal tail in the nanomolar range (for example, Cpr6 has a K_D of 240 nM (Prodromou *et al.*, 1999)). The activation of kinases, which are clients of Hsp90 (reviewed in Pearl, 2004), requires the activity of the co-chaperone Cdc37, which is itself regulated by the phosphatase activity and simultaneous interaction of the TPR co-chaperone PP5 with the C-terminus of Hsp90 (Vaughan *et al*., 2008). Hsp90 displaces the autoinhibitory domain from its interaction with the catalytic domain (Yang *et al*., 2005), and in this way acts both as the negative and positive regulatory hub for kinase activation *in vivo*. The TPR co-chaperone Hop (Sti1 in yeast) also acts as an adaptor and is a functional link between the mechanisms of the Hsp70 chaperone machine and Hsp90

(Onuoha *et al*., 2008). Hop contains three TPR domains that show different affinities for the C-terminal EEVD motif, which is also present in Hsp70 (Figure 1.10). An interesting feature of this interaction is that Hop shows a preference for the ADP-bound form of Hsp70 (Meimaridou *et al.*, 2009) and the nucleotide-free open conformation of Hsp90 (Onuoha *et al*., 2008), which exposes the largest proportion of hydrophobic patches within its V-shaped cleft. Hop has been linked to prion diseases (Zanata *et al*., 2002) and can also promote the folding of the non-native Hsp90 substrate luciferase (Johnson *et al*., 1998). By preferentially binding the open conformation, Hop provides the largest possible folding landscape and 'blank slate' for polypeptides being shuttled from Hsp70 to the late-stage folding action of Hsp90.

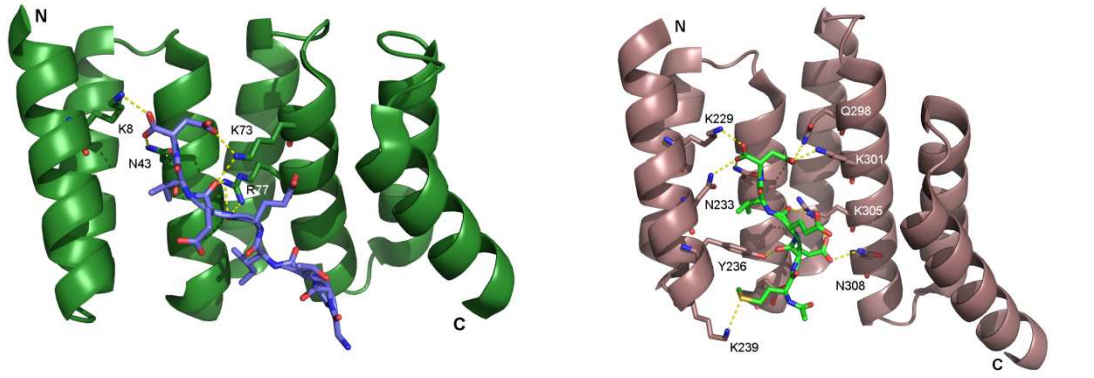


Figure 1.10 **Crystal Structures of the TPR1 and TPR2A Domains of Hop/Sti1 in Complex with Chaperone Peptides** (PDB IDs: 1ELR and 1ELW) – Two of the alpha-helical TPR domains of Hop1 interact specifically with peptides derived from HSP70 (**Left**) and Hsp90 (**Right**). The central interaction herewith is a hydrogen bond network (shown as **dashed lines**) between a lysine and two asparagine residues to the aspartate residue in the chaperone peptides. The Hsp90 peptides makes more significant contacts with the TPR motif, with six residues of TPR2A involved in the interaction with the MEEVD peptide (**Right**), as opposed to four from TPR1 with the EEVD Hsc70 peptide (**Left**).

Sgt1 is a TPR-containing co-chaperone of Hsp90 with a multi-domain architecture. While it has no measureable affect on the ATPase activity, it displays a preference for the ADP-bound lid-open conformation (Catlett *et al.*, 2006). Given that the N-terminal dimer is incompatible with Sgt1 binding (Zhang *et al*., 2008), the Hsp90:Sgt1 interaction depends partly on structural conformations of Hsp90.

1.2 THE HSP90 CO-CHAPERONE, SGT1

Sgt1 was originally identified as a dosage suppressor of *skp1-4*, a mutant of the inner kinetochore protein Skp1, which exhibits a chromosome missegregation phenotype (Kitagawa *et al.*, 1999). Sgt1 was found to be essential for progression through both the G1/S and G2/M transitions in the cell cycle, with the latter dependent on the ability of Sgt1 to chaperone both Skp1 and Ctf13 (the structural core of the inner kinetochore (Russell *et al.*, 1999)) to the centromere. This process is regulated by MAD2, a spindle checkpoint protein in yeast (Gardner *et al.*, 2001). Mutant *sgt1-3* cells were found to be defective in the assembly of the CBF3 complex on CDEIII DNA, due to the lack of activation of Ctf13 (Kitagawa *et al.*, 1999). The authors concluded that *sgt1-3*, which carries a double mutation in the N-terminal region of Sgt1, is unable to form complexes with Skp1, which is itself required for the activation of Ctf13, the formation of the structural heterodimer Skp1:Ctf13 of CBF3 and consequently, the anchoring of the kinetochore on centromeric DNA.

Sgt1 fulfils numerous roles in different organisms. In fruit flies, for example, Sgt1 was recently shown to be critical for specifying cellular morphogenesis and polarity (Andersen *et al.*, 2012; Eastburn *et al.*, 2012), while in yeast, Sgt1, besides its role in kinetochore assembly, has an additional function in protein degradation (Kitagawa *et al.*, 1999) and in cyclic AMP (cAMP) signalling (Dubacq *et al.*, 2002; Flom *et al.*, 2012). One of the most prominently characterised roles of Sgt1 is in the context of plant immunity and the host immune response (reviewed in Kadota *et al.*, 2010). Here, Sgt1 performs a role in maintaining the steady-state levels of Resistance (R) genes (Azevedo *et al.*, 2006) and the activation of nucleotide-binding leucine rich repeat (NB-LRR) type receptors in the host immune response (Schulze-Lefert *et al.*, 2004),

The finding that Sgt1 also associates with Hsp70 supports a role for Sgt1 as an adaptor protein (Spiechowicz *et al.*, 2007), which suggests that it might perform a similar function to Hop in promoting the folding and correct presentation of binding sites that potentiate downstream signalling within these pathways.

1.2.1 INSIGHTS INTO THE STRUCTURE OF SGT1

Sgt1 has a modular architecture consisting of three domains: an N-terminal helical tetratricopeptide repeat (TPR) domain, a 'CHORD-containing and Sgt1' domain, with a β-sandwich fold found in the co-chaperone p23 and Hsp20/α-crystallin molecules, and a putatively unstructured 'Sgt1 specific' SGS domain at the extreme C-terminus (Figure 1.11).

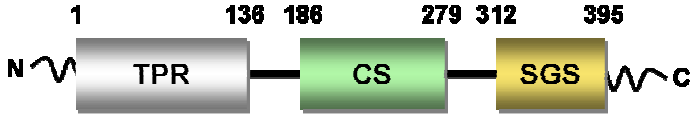


Figure 1.11: **The Domain Arrangement of Yeast (*S. cerevisiae*) Sgt1** – Sgt1 consists of three separate domains interspersed by linkers that are speculated to be flexible and aid in the function of Sgt1. The N-terminal TPR domain (residues 1-136) is followed by the CS domain (residues 186-279) and the SGS domain at the C-terminus (residues 312-395). The linker between the CS and SGS domain, as well as the region between the TPR and CS domain, are thought to have regulatory functions in dimerisation (see discussion) (Bansal *et al.*, 2008; Bansal *et al.*, 2009).

While these domains have been classified as part of the overall architecture of Sgt1, their boundaries are less well defined, with the current knowledge suggesting that they are interspersed by linker regions of unknown function.

Studies of Sgt1 mutants have shown that at least its N-terminal TPR and C-terminal SGS domains contribute in different cellular contexts, and that these domains specify Sgt1 for diverse functions in different species. At one end, the TPR domain mediates interactions with the Skp1:Ctf13 heterodimer through the BTB/POZ domain of Skp1 (Bansal *et al.*, 2004; Catlett *et al.*, 2006) in yeast, and in plants, the physical link of the TPR and SGS domains provides cross-talk between pathways of Resistance genes and NB-LRR proteins, respectively (Azevedo *et al.*, 2006; Schulze-Lefert *et al.*, 2004). In this context, binding of NB-LRRs to the SGS domain promotes their maturation to prime the plant immune response (Austin *et al.*, 2002). Mutations in the SGS domain also effect Cyr1p binding, an interaction important in cAMP signalling (Dubacq *et al.*, 2002). The cooperativity of the TPR and SGS domains in plant immunity shows that Sgt1 can link together

two related pathways, which might be important in the context of kinetochore assembly, where the TPR domain binds the CBF3 component Skp1. While it has not been shown directly, binding of the LRR-containing kinetochore component Ctf13 to the SGS domain would link together two components of the inner kinetochore via a single structural scaffold.

Sgt1 also interacts with the molecular chaperone machine via its CS domain (Bansal *et al.*, 2004; Lee *et al.*, 2004), which lies in between the TPR and SGS domains. To date, this region has not been found to engage in functions other than guiding the molecular chaperone machinery to the cellular processes facilitated by the N- and C-terminal domains of Sgt1. Detailed, atomic resolution structural information for Sgt1 is currently limited to this region. Two crystal structures have been solved to-date, one of the Sgt1 CS domain bound to the N-terminal ATPase domain of Hsp90 (PDB ID: 2XCM; Figure 1.14, A) (Zhang *et al.*, 2008), and another showing Hsp90 interacting simultaneously with the Sgt1 CS domain and an additional co-chaperone Rar1 (Zhang *et al.*, 2010). The larger complex highlights the cooperative binding of two co-chaperones to Hsp90, which might have some functional relevance when considering the plethora of roles of Sgt1. Even though the N-terminal TPR domain was initially suspected of binding Hsp90, the CS domain was found to be sufficient for this interaction (Lee *et al.*, 2004).

1.2.1.1 Structural and Functional Studies of CS domain-containing Proteins

A CS domain has a seven strand β-sandwich type fold, with one four-strand and one three-strand face. The strands making up both sides are interwoven in sequence, forming a highly compact and stable overall fold (Figure 1.12).

The CS domain fold is a common protein:protein interaction motif in eukaryotes and, in this respect, is not unique to Sgt1. Many of these have been found to bind Hsp90, while others do not. Shq1, a protein involved in processing of ribosomal RNA, shows no interaction with Hsp90 in the presence and absence of nucleotide in solution NMR experiments (Singh *et al.*, 2009). Shq1 itself shares little

sequence similarity with the Hsp90 co-chaperone and canonical CS-containing protein p23 or Sgt1, and replaces a beta-strand in p23 with an alpha helix, which is a feature also lacking in Sgt1. The CS domain of the Siah1-interacting protein, SIP, is critical in the degradation of beta-catenin by binding the E3 ligase Siah-1 in this pathway (Bhattacharya *et al.*, 2005). While it shares some functional overlap with Sgt1 in binding Skp1 (and also contains a C-terminal SGS-like domain), it does not undertake its native function in the context of Hsp90. SIP, like Shq1, has additional features that are closely juxtaposed to its CS domain (PDB ID: 1X5M). This suggests that other domains might play a role in modulating Hsp90 binding to CS domains, in turn increasing the diversity of roles for these types of proteins.

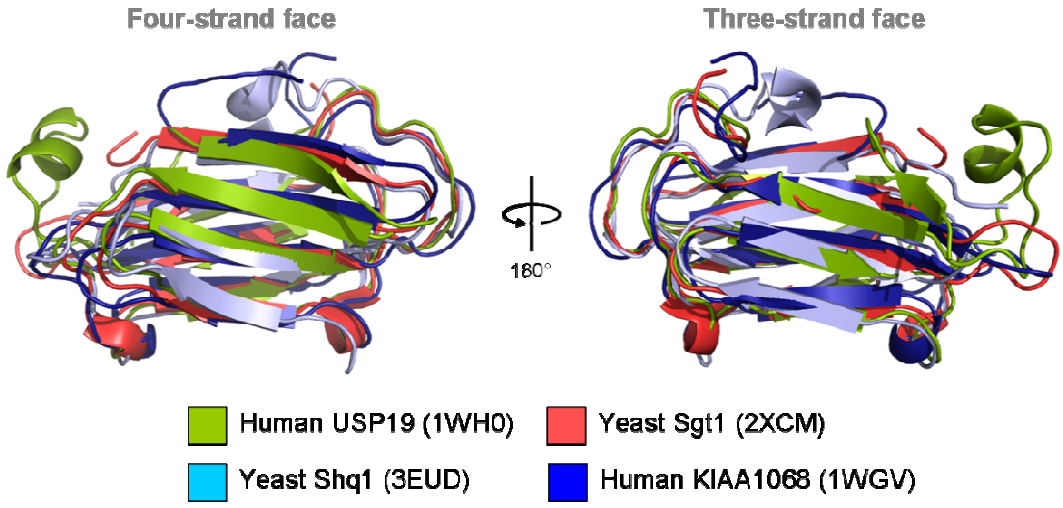


Figure 1.12: **Structural Superposition of CS Domains from Different Proteins** – CS domains exhibit a β-sandwich with a four- and a three-strand face that are related by 180°. Superimposing the CS domains from different proteins shows that the overall fold is highly conserved in structure.

Those CS domains that bind Hsp90 show little consistency in their preference for a particular nucleotide-bound state of the chaperone. In the co-chaperone p23/Sba1, the CS domain stabilises the lid-closed ATP-bound form of Hsp90 and limits the basal rate of ATP hydrolysis until client protein binding (McLaughlin *et al.*, 2006). Despite sharing a very similar fold to p23 (Figure 1.13) (Ali *et* al., 2006; Zhang *et al.*, 2008), Sgt1 conversely shows a preference for the ADP-bound lid-open conformation of the Hsp90 N-terminus and does not influence its ATPase activity (Catlett *et al.*, 2006), but requires its entire N-terminal domain to

mediate this binding *in vivo* (Bansal *et al.*, 2004). Similarly, the interaction of the co-chaperone Chp1, involves its CS domain and support from the neighbouring CHORD-II domain, itself a CS domain-like fold (Wu J *et al.*, 2005). In contrast to both Sgt1 and p23, Chp1 interacts with Hsp90 independently of nucleotide-driven changes to the lid-segment. These inconsistent preferences infer that CS domain-containing proteins might perform a more general role as adaptors for client proteins of Hsp90. Furthermore, the fact that Hsp90-binding CS domains also show a dependence on additional domains suggests that binding specificity is determined at the sequence level.

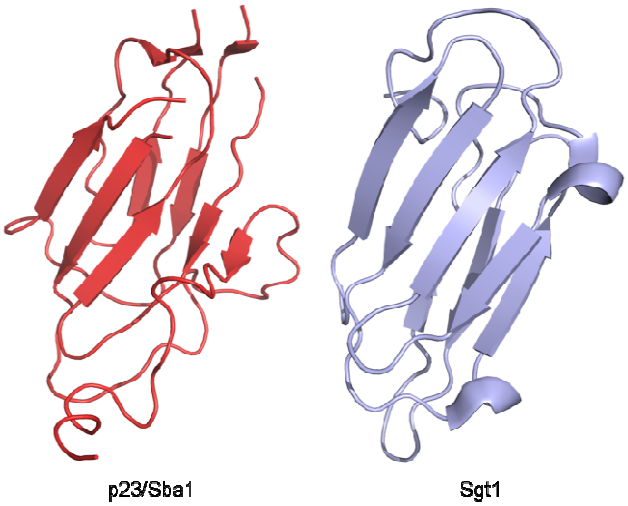


Figure 1.13: **Comparison of the Structure of the CS Domain of Sgt1 and the p23/Sba1 Hsp90 Co-chaperone** (PDB IDs: 2CG9 and 2XCM) [Image adapted from (Shirasu, 2009)] – The CS domain and p23 share an overall structure, with an open beta-sheet face making up the principal interacting interface with Hsp90. However the two proteins exhibit very distinct binding modes to Hsp90.

In Sgt1, interactions with Hsp90 are mediated via the four-strand face of the CS domain, with residues being contributed primarily from β1 (Figure 1.14, B and C). This region of Sgt1 is highly conserved amongst all homologues, including yeast Sgt1 (Zhang *et al.*, 2008), meaning that this interaction is likely to be preserved in different organisms. Sgt1 can form quaternary complexes with proteins such as RAR1, Hop/Sti1 and Skp1 (Bansal *et al.*, 2004; Zhang *et al.* 2010), implying that it acts cooperatively within the biological contexts of these proteins. Deletion studies affecting the Sgt1 CS domain first provided evidence that Hsp90 forms an additional interface with sequences C-terminal to the CS domain (i.e. the SGS domain) (Spiechowicz *et al.*, 2007), however based on the crystal structure, it is clear that the deletion of residues C-terminal to Glu262 (present at the centre of β7) would disrupt the fold of the CS domain and abolish

the interaction with Hsp90 (Figure 1.14). It is likely that the SGS domain therefore works exclusively with Hsp70 and LRRs, and not Hsp90.

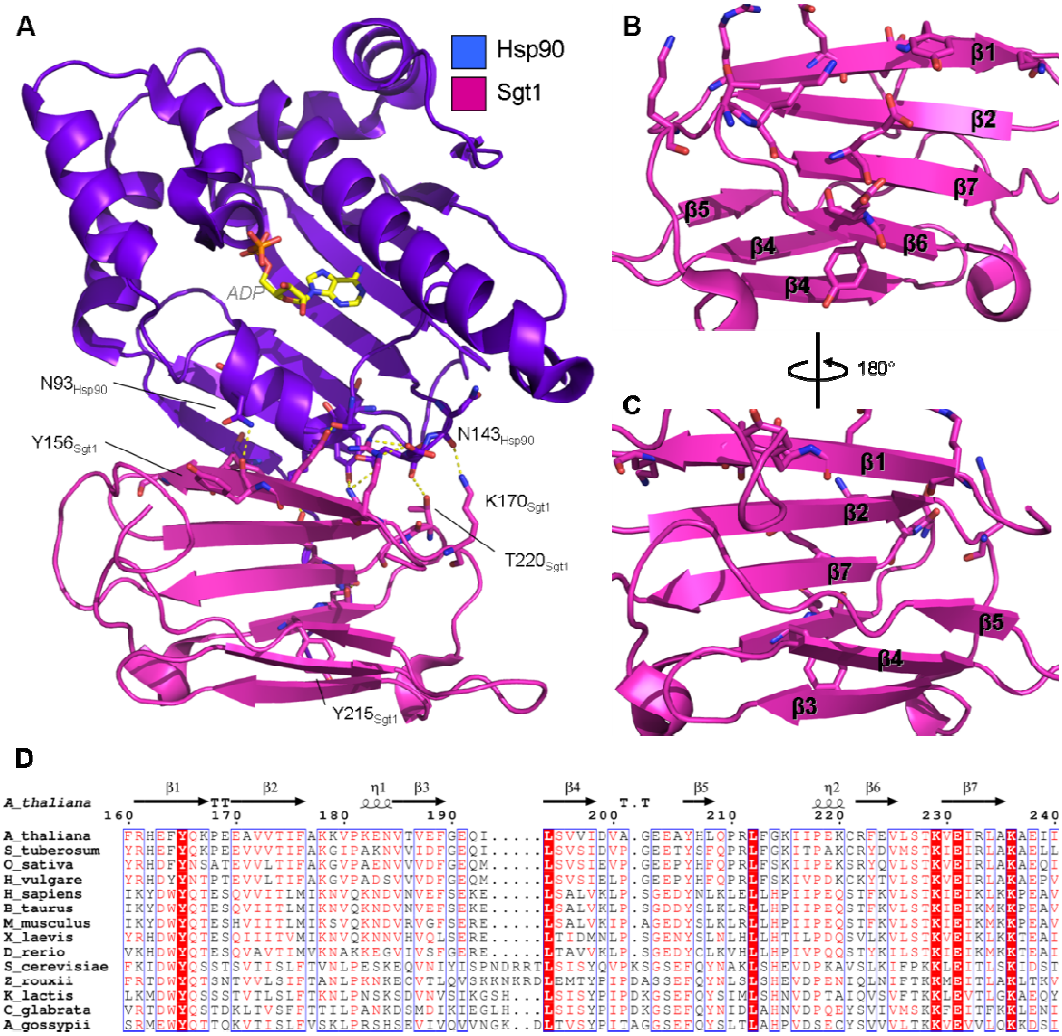


Figure 1.14: **Structure of the Sgt1 CS Domain in Complex with the N-terminal ATPase Domain of Hsp90** (PDB ID: 2XCM) – **(A)** Crystal structure of the CS domain of Sgt1 (**Magenta**) in complex with the ADP-bound N-terminus of Hsp90 (**Slate**). The core of the interaction occurs across the four-strand face of the CS domain (**Magenta**), with residues available in the lid-open conformation of Hsp90 leading to preferential binding of this conformational state of the chaperone to Sgt1. **(B)** View of the four-strand face of the Sgt1 CS domain in isolation, with the different strands of the domain labeled. Despite the uniform nature of this domain, the strands encoded in sequence are interwoven, forming an intricate and tight β-sandwich structure. **(C)** View of the reverse side of the CS domain, with the three-strand face in the foreground. **(D)** Sequence alignment of *Arabidopsis thaliana* Sgt1a (visible in the crystal structure 2JKI) and other Sgt1 homologues showing the conservation of residues within the CS domain, with the secondary structure of the CS domain imposed above the sequence. The image was generated using ESPript (Gouet *et al.*, 1999).

1.2.1.2 STRUCTURAL AND FUNCTIONAL STUDIES OF TPR DOMAIN-CONTAINING PROTEINS

Our understanding of the domain family to which the N-terminus of Sgt1 belongs benefits greatly from its ubiquity as a protein:protein interaction module. The family of tetratricopeptide repeat (TPR) motif-containing proteins are common to a set of co-chaperones and more generally comprise a set of protein:protein interaction modules that span a variety of different cellular processes. These include the regulation of mitochondrial morphology (Suzuki *et al*., 2003), regulation of glucocorticoid receptors and potassium channels (Chinkers, 2001), and signal recognition via the TPR-containing membrane-associated protein Tom20 (Abe *et al*., 2000). In co-chaperones of Hsp90, TPR domains typically interact with the C-terminal MEEVD motif via an interaction with the concave groove formed by the TPR (see Section 1.1.3). Since Sgt1 interacts via its CS domain with Hsp90, it represents an exception to this common behaviour of TPR-containing Hsp90 co-chaperones.

A single TPR comprises 34 amino acids that form a helix-turn-helix motif, where the two helices pack together in a slightly twisted conformation. The core packing is mediated by a pattern of large (positions 4, 7, 11 and 24) and small (positions 8, 20 and 27) hydrophobic residues rather than a specific amino acid (Figure 1.15, B). Helix-breaking residues dictating both inter- and intra-TPR turns are also conserved. While some positions are more tolerant of mutations, others are relevant in defining the cellular function of a particular TPR and therefore tend to be conserved in homologues (D'Andrea *et al.., 2*003, Main *et al.., 2*003). Some TPR motifs exist in isolation, flanked by structural features that have an effect on the function of the TPR (such as Tom20), though most TPR motifs exist in multiple helical arrays, forming a protein:protein interaction module with a right-handed superhelical arrangement (Figure 1.15, A). Sgt1, for example, forms a three-repeat TPR domain. The number of repeats and their amino acid composition defines context, function and regulation of the TPR-containing protein. In addition, TPR domains are often capped by a hydrophilic C-terminal helix (called the 'capping helix'), which influences the solubility of the TPR-containing protein (Main *et al*., 2003).

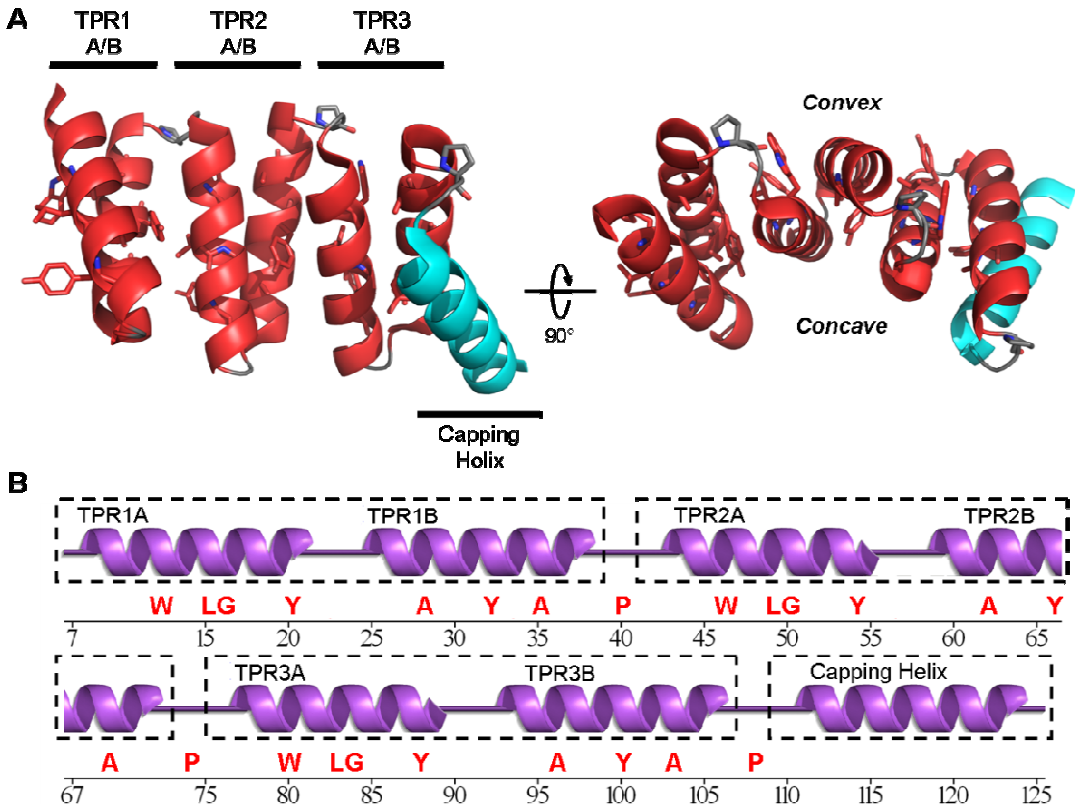


Figure 1.15: **Cartoon representation and structure of a consensus three-repeat TPR** (PDB ID: 1NA0) – **(A)** A tetratricopeptide repeat is a helical motif of a pair of helices, which typically pack together in multiples to form a curved structure with a superhelical twist. To the left is a simplified cartoon depiction of the TPR, highlighting the helical pairs (1A/B to 3/AB) in **Red** and the C-terminal capping helix, a feature commonly found in TPR domain proteins, in **Light Blue**. The model depicts the arrangements of the individual TPRs as they exist in the structure of CTPR3, an idealised TPR protein representative of three-repeat TPRs (PDB ID: 1NA0). As a result of its fold, the TPR bears convex and concave faces (**Right**). **(B)** Linearised secondary structure of 1NA0. The consensus residues preserved within the primary sequence, and critical to the tertiary fold of the TPR, are listed in **Bold Red Capital** lettering beneath the graphical depiction of the secondary structural elements. TPR helices are labeled above these features. This figure was adapted from the output of PDBsum (Laskowski, 2009).

While some TPR-containing proteins are monomeric, a number of them – including Sgt1 (Bansal *et al.*, 2008) – are known to self-associate via their TPR domain (reviewed in Zeytuni *et al.*, 2012). Studies of an idealised three-repeat TPR domain with three tyrosine point mutations (CTPR3Y3) exemplify the structural importance of self-association for function (Figure 1.16). Oligomers of CTPR3Y3 that form via the inter-TPR loop regions depend on the presence of tyrosines in these regions between TPR repeats (Krachler *et al.,* 2010). This type of interaction results in the stacking of opposing protomers (Figure 1.16, 'stacking'), and is stabilised by peripheral hydrogen bonds from residues available across this interface. The yeast mitochonrdial membrane protein Fis1, involved in mitochondrial fission, exhibits this mode of self-association *in vivo*,

albeit in a manner that results in an intra-TPR loop stacking interaction (Zhang Y *et al., 2*007). Krachler *et al.* also found that displacement of the capping helix allowed their designed TPR domain (CTPR3Y3) to self-associate end-on with another protomer (Figure 1.16, 'end-on'). In this interaction, two adjacent TPR protomers form an elongated superhelical twist through the packing of adjacent TPR3 helices.

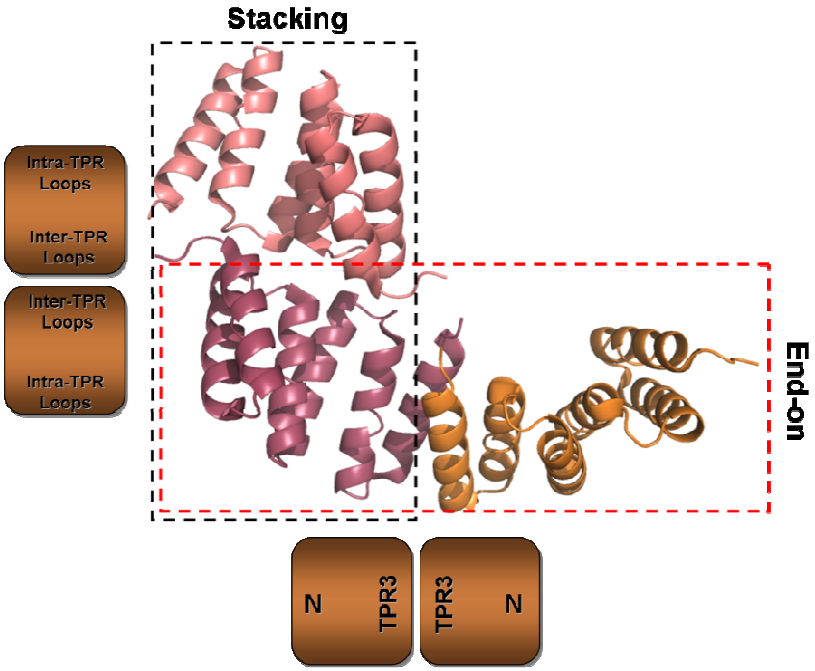


Figure 1.16: **Different modes of self-association of TPR protomers** (PDB ID: 2WQH) – The consensus three-repeat TPR protomer CTPR3Y3 (in **Cartoon** representation) highlights two potential modes of self-association of TPR-containing proteins, as determined by Krachler *et al.*. In their study they determined that an idealised TPR could form a stacking mode of self-association mediated by the loops of the TPR domain (**Side**) or a hydrophobic, end-on association (**Bottom**), for which displacement of the C-terminal capping helix is critical. Fis1 represents an example of a protein that exhibits a stacking-type interaction *in vivo*.

As a result of both the stacking and end-on interfaces, the effective binding surface of a single TPR protomer changes. In this manner, self-association could specify sites for ligand binding in naturally occurring TPR domain-containing proteins. For example, in self-association of the *Shigella* type III secretion chaperone IpgC, dimerisation produces back-to-back asymmetric packing of TPR domains, involving a swap of helices occurring before its core three-repeat

domain (Lokareddy *et* al., 2010). This interaction yields interfaces that mediate binding of a single copy of the epithelial cell invasion mediator IpaB.

In the mitotic checkpoint protein Bub1, where self-association occurs via a stacking-type loop-mediated interface, dimerisation is crucial to its interaction with the kinetochore protein blinkin (Bolanos-Garcia *et al*., 2009). The TPR domain was found to be crucial in mitigating aberrant chromosome alignment in humans and the Bub1:blinkin interaction important for localising Bub1 to the kinetochore to initiate the mitotic checkpoint (Kiyomitsu *et al*., 2007).

TPR domains can recognise a wide variety of cellular ligands with a high degree of specificity. Depending on the structural landscape of the binding site, ligands will bind in either an extended or helical conformation, or both (Zeytuni *et al*., 2012). The binding of Hsp90 and Hsp70 peptides to Hop, for example, shows that the C-terminal EEVD motif binds in an extended conformation (Onuoha *et al*., 2008). The contacts that are made by the extended peptide maximise the number of hydrogen bonds that are formed in the interaction and serve, in turn, as a means of specifically recognising Hsp70 and Hsp90 ligands by TPR1 and TPR2A of Hop, respectively.

Ligands may also bind in a helical conformation. In the mitochondrial import protein Tom20, the TPR domain recognises signal peptides in an amphipathic helix conformation. Leucines that are aligned on one side of the helix are projected into the hydrophobic core of the concave groove and in this manner, facilitate specific recognition of the mitochondrial import signal (Abe *et al*., 2000). In the case of the aforementioned yeast mitochondrial membrane protein Fis1, a 'helical clamp' that forms over the concave and convex binding surfaces of the Fis1 TPR domain by the adaptor proteins Mdv1 and Caf4, represents another example of helical recognition by a TPR-containing protein (Zhang Y *et al*., 2007). This interaction is critical to the recruitment of the dynamin-related GTPase Dmn1 to the mitochonrdial membrane to promote membrane fission. These examples emphasise the ability of TPR domains to specifically recognise a large number of ligands in the cellular milieu.

TPR domains perform a common role in scaffolding protein:protein interactions, but despite this, share little sequence identity with each other outside of the core TPR (Allan *et al.*, 2011). By virtue of their evolution as helical repeat proteins, TPR domains have become versatile units in biology, which perhaps reinforces the impact of sequence delineation within non-consensus regions (Krachler *et al.*, 2010). The manner in which TPR domains facilitate downstream interactions and recognise their cognate binding partners is key to understanding the widespread use of these repeat-based protein recognition motifs in eukaryotes and prokaryotes. In this regard, obtaining the structure of the N-terminal TPR domain of Sgt1 will highlight its structural contribution to functions relevant in the assembly of the kinetochore.

1.2.2 The Role of Sgt1 in the Assembly of the Yeast Kinetochore

There are two known classes of Sgt1 mutants affected at different stages of the cell cycle. G1 alleles of Sgt1 arrest at the G1/S checkpoint, and are believed to reflect the role of Sgt1 in protein turnover via Skp1 and SCF ubiquitin ligase complexes. G2 alleles are affected at the G2/M transition and show a chromosome missegregation phenotype (Kitagawa *et al.*, 1999), a phenotype mirrored in mammalian cell lines when Sgt1 is disrupted by RNA interference (Steensgaard *et al.*, 2004). This classification of Sgt1 mutants re-emphasises the independent and modular nature of its individual domains, with G1 alleles affected predominantly in the SGS domain (believed to mediate interactions with LRRs, a critical component of SCF complexes (see Section 1.3.1.3)) and G2 alleles impaired in their ability to bind Hsp90 and Skp1. Since Skp1 links Sgt1 to the kinetochore, it is unsurprising that mutants in the TPR region, which binds Skp1 (Bansal *et al.*, 2004), are arrested at the G2/M checkpoint. This is further supported by the interaction of Skp1 with Bub1 (BubR1 in humans), a spindle-assembly checkpoint protein that monitors the level of tension present at the DNA-microtubule bridge (Kitagawa *et al.*, 2003). In humans, BubR1 is not recruited to kinetochores in Sgt1-depleted cells, suggesting that checkpoint assembly is dependent on a function of Sgt1 (Steensgaard *et al.*, 2004). Given the parallel roles of Sgt1 in both humans and yeast, it is therefore not possible to

determine from the existing literature whether the phenotype of G2 alleles of Sgt1 results from the lack of a Bub1 recruitment to the kinetochore or from the disruption of a direct Sgt1:CBF3 interaction leading to failed assembly of the kinetochore. Nevertheless, *sgt1-3* was shown to be specifically impaired in CBF3 formation (Kitagawa *et al.*, 1999), suggesting that at least a subset of these mutants is relevant in the regulation of the latter. *Sgt1-3* is one of two mutants of the four G2 alleles affected in both the core TPR (L31P and F99L) and the CS domain (N213L), with the other being *sgt1-7* (L117P and Q191K). Table 1.1 provides a summary of all mutant phenotypes of yeast Sgt1 characterised to-date (Cherry *et al.*, 2012; Bansal *et al.*, 2004).

Mutant	Arrest	Mutations	Domains Affected by Mutations
sgt1-3	G2/M	L31P, F99L, N213L	TPR, CS
sgt1-6	G2/M	L26P, F99L	TPR
sgt1-7	G2/M	L117P, Q191K	TPR, CS
sgt1-12	G2/M	A74T, I151V	TPR, TPR-CS linker
sgt1-1	G1/S	Y119C, S378P	SGS
sgt1-2	G1/S	L243H, S378F	CS, SGS
sgt1-4	G1/S	K360E, S385T	SGS
sgt1-5	G1/S	D220V, E364K	CS, SGS
sgt1-8	G1/S	K360E	SGS
sgt1-9	G1/S	K360E	SGS
sgt1-10	G1/S	S378P	SGS
sgt1-11	G1/S	S378P, M390V	SGS

Table 1.1: **Summary of Isolated Sgt1 Mutants** – Details of point mutants of Sgt1 utilised in numerous studies of its function *in vivo*, derived from the Saccharomyces Genome Database (Cherry *et al.*, 2012) and (Bansal *et al.*, 2004). These mutations map to various domains of Sgt1 (**Far Right Column**) underlining the different functions mediated by these domains *in vivo*.

Direct evidence for a role of the Hsp90:Sgt1 chaperone machinery in the initiation of kinetochore assembly comes from its active role in assembling the structural core of the CBF3 complex in the form of the Skp1:Ctf13 heterodimer (Bansal *et al.*, 2004; Stemmann *et al.*, 1996). Critical to this function appears to be the dimerisation of the TPR domain (Bansal *et al.*, 2008), a feature that is differentially conserved between homologues of Sgt1 (Nyarko *et al.*, 2007), despite the overall conservation of its domain architecture across species. Yeast (*S. cerevisiae*), barley (*Hordeum vulgare*) and water cress (*Arabidopsis thaliana*) Sgt1 all show concentration-dependent dimerisation in size exclusion chromatography experiments, while human Sgt1 does not. Similar experiments

on the TPR domain of barley Sgt1 alone showed that this domain was responsible for mediating self-association, which itself is mediated by an intra-TPR disulphide bond (Nyarko *et al.*, 2007). In the context of plant immunity, this disulphide bond might function as a redox sensor for the oxidative burst spurned by an invading pathogen (Spoel *et al.*, 2011).

Immunoprecipitation experiments first showed that Sgt1 could form dimers *in vivo* (Bansal *et al.*, 2008). Truncation of Sgt1 to Gln211 was still permissive of homodimerisation between two identical constructs, but further truncation to Ser182 reversed this behaviour *in vitro*. In addition, the same truncation was unable to bind to Skp1 when complexed with full length Sgt1, showing a critical structural role for the full length Sgt1 dimer in its interaction with Skp1. The mutation of *sgt1-3* affects the core TPR domain at Leu31 and is unable to form dimers, leading to decreased viability at non-permissive temperatures. The overexpression of a chimera of *sgt1-3* and the dimerisation domain of CENP-B rescues yeast grown at non-permissive temperatures (Bansal *et al.*, 2008). Together this data stresses the importance of a dimer of Sgt1 for mitigating the chromosome missegregation phenotype.

Phosphorylation of the C-terminal SGS domain disrupts dimerisation (Bansal *et al.*, 2009). This modification occurs at Ser361 within the consensus site for casein kinase 2 (CK2) (Pinna, 1990) and a phosphomimetic (S361D) mutant expressed endogenously proved lethal and unable to form CBF3 complexes *in vivo*. In contrast, phosphorylation of Ser331 by Polo-like kinase 1 (Plk1) in the human homologue of Sgt1, enhances the association of Sgt1 and Hsp90 *in vivo* (Liu *et al.*, 2012), and promotes the stabilisation of the essential kinetochore complex Mis12, which is essential to the formation of kinetochore-microtubule attachments (Davies *et al.*, 2010). Previous *in vitro* studies of Sgt1 have shown that the SGS domain alone does not bind the TPR domain (Bansal *et al.*, 2004), however it is plausible that phosphorylation stimulates its reordering or permits further tertiary interactions within the domain that are transmitted to the N-terminus, a common structural consequence of phosphorylation (reviewed in Johnson *et al.*, 2001). In the absence of structural information on both the

phosphorylated and unphosphorylated form of Sgt1, it is difficult to gauge its precise effect on the TPR domain.

The specific function of Hsp90 in kinetochore assembly remains unresolved, though the fact that CK2 is a well-characterised client (reviewed in Miyata, 2009), and the stability of Polo kinase in *Drosophila* flies is also dependent on Hsp90 (de Cárcer *et al*., 2001), suggests that it might be related to the phosphorylation of Sgt1. Numerous kinases are known to associate with Hsp90 as clients, with the chaperone machinery allowing their maturation and activation (reviewed in Pearl, 2005). Regulation of the assembly of the inner kinetochore complex, CBF3, is thought to be occur through the phosphorylation of Ctf13 (Kaplan *et al.*, 1997), the structural core of this complex (Russell *et al*., 1999). Whether through folding, or bridging and transient activation of CBF3 components, it is clear that Hsp90 is involved in some capacity in the initiation of kinetochore assembly, given the sensitivity of Sgt1 mutants to the Hsp90 inhibitor geldanamycin (Bansal *et al.*, 2004).

1.3 THE YEAST KINETOCHORE

The eukaryotic kinetochore is a proteinaceous assembly paramount to the fidelity of chromosome segregation and progression of the cell cycle. In the process of cell division, the kinetochore fulfills two holistic functions, acting as a functional hub for signalling and motor proteins which aid in chromosome segregation, and a structural link between sister chromatids and microtubules. The complexity of the kinetochore suggests that its assembly must be tightly controlled in a manner that enforces its localisation to the centromere and its role as a molecular bridge between DNA and the mitotic spindle apparatus. Several studies over the past decades have unveiled in great detail the biological functions of the yeast *Saccharomyces cerevisiae* kinetochore, and it has therefore become a model organism for this area of study (Welburn and Cheeseman, 2008).

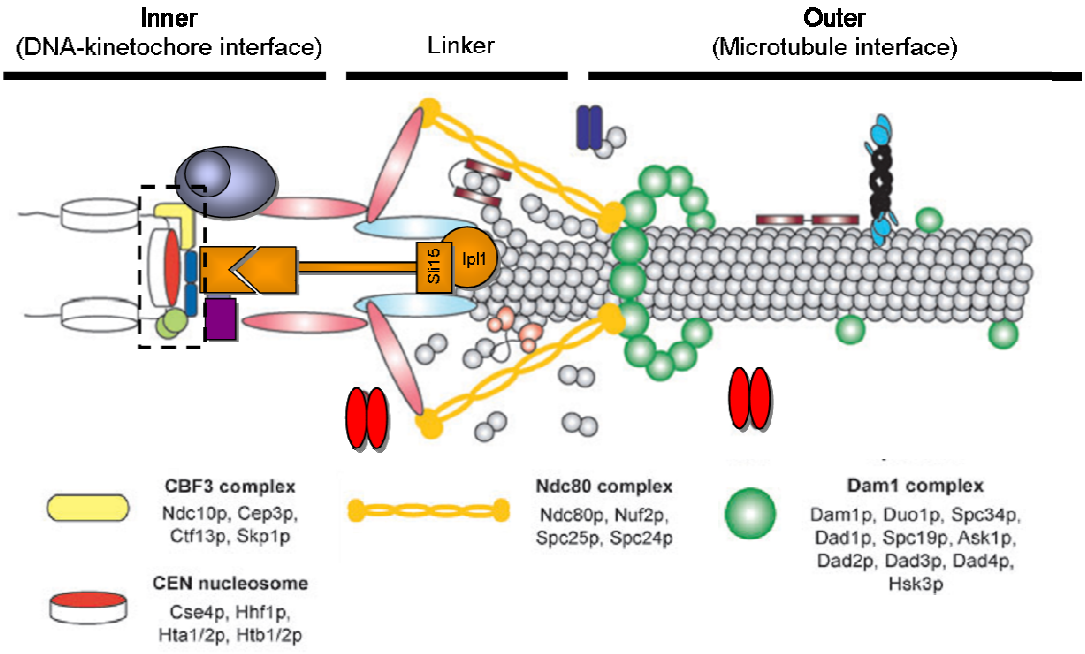


Figure 1.17: **Overall Architecture of the Yeast Kinetochore** [Image adapted from (Westermann *et al.*, 2007)] – The above diagram summarises the complex layering of the yeast kinetochore. The inner layer forms an intimate interaction with centromeric DNA, while the outer layers are assembled in a fashion conducive to the processive movement of polymerising microtubules. Though it is not evident from this diagram, the kinetochore fulfils the functions of being an anchor point for microtubules and a recruiter of motor proteins. The **Boxed** region on the left-hand side of the diagram highlights the inner kinetochore and the area of interest of this study, the CBF3 complex (**Yellow shape**).

The kinetochore of *S. cerevisiae* can be divided into three layers: an inner anchoring layer, central bridging complexes (the 'linker' layer) and an outer microtubule interface. The linker and outermost layer are incredibly complex, consisting of diverse families of proteins (Figure 1.17), with proteins in the outer layer primarily responsible for maintaining the tension imposed on the kinetochore structure by the force of microtubules during the segregation of sister chromatids (Westermann *et al.*, 2007). However, it is the complexes of the linker layer that dynamically monitor the correct orientation of kinetochores before and during alignment on the metaphase plate.

The biorientation (i.e. opposing orientation) of kinetochores is important in the shuttling of sister chromatids to opposite poles of the cell. The linker layer, which forms a rough scaffold between the inner and outer aspects of the kinetochore complex, is important in recruiting proteins such as Mad2 and Bub1 to initiate the spindle assembly checkpoint (SAC) (Westermann *et al.*, 2007). In higher

eukaryotes, a critical mass of SAC components persists during the attachment of microtubules to ensure the biorientation of kinetochores at the metaphase plate and maintain the necessary tension for kinetochore-DNA transport (Hoffmann *et al.,* 2001). The prototypical linker complex, Ndc80, forms a 570 Å coiled-coil structure with two globular heads (Figure 1.17) and recruits the Mad1-Mad2 checkpoint complex and the molecular motor dynein (Sironi *et al.*, 2002; McLeland *et al.,* 2003). Mutations to Nuf2 (part of the Ndc80 complex) disrupt the kinetochore-microtubule interaction and result either in the loss of cell viability or the engagement of the SAC (He *et al*, 2001). As part of their sensory function, the linker proteins also play a role in recruiting kinases to modulate kinetochore-microtubule attachment. The conserved kinase Ipl1 (Aurora B in humans) phosphorylates Dam1 (Keating *et al.*, 2009) and Ndc10 (Biggins *et al.,* 1999), in the outer and inner layers respectively, and forms a complex with Sli15 within the linker layer (Tanaka *et al.*, 2002). In its role with Sli15, Ipl1 transiently destabilises the kinetochore-microtubule attachment of monotelic or syntelic configurations that do not hold the necessary tension to facilitate proper segregation, in order to permit a second opportunity for attachment in the correct bioriented configuration. Only when proper tension is present at the DNA-microtubule bridge does the checkpoint permit the continuation of the cell cycle (Tanaka *et al.*, 2002). In this manner, components of the linker layer perform a regulatory role in the kinetochore, sensing the signals of checkpoint proteins, and propagating these signals to motor proteins, the spindle poles and the DNA anchor.

The role of the kinetochore as the DNA-microtubule bridge is ultimately dependent upon its ability to bind DNA, and for these DNA-binding components to form the basis for hierarchical assembly of the linker and outer layers. This occurs via the heteromeric CBF3 complex (centromere binding factor 3) (Figure 1.17, Dashed Box), which consists of a heterodimer of Skp1:Ctf13, and homodimers of Cep3 and Ndc10 (Espelin *et al.*, 1997), with the latter two forming the primary DNA binding surface of the kinetochore. High salt size exclusion chromatography experiments emphasised the stability of a complex of Cep3 and the Skp1:Ctf13 heterodimer, and an oligomeric complex of Ndc10

(Stemmann *et al.*, 1996). The formation of CBF3-like complexes is limited to centromeres of a 'CDEI-CDEII-CDEIII' (centromere-determining elements 1-3) composition found exclusively in yeast bearing point-centromeres (Meraldi *et al.*, 2006). The equivalent complex in humans, CCAN (constitutive centromere associated network) is constitutively associated with centromeric DNA and at least epigenetically specifies the localisation of the kinetochore to the parent strand of DNA (Hori *et al.*, 2008). Whether or not CBF3 is constitutively associated with the centromere throughout the cell cycle is currently unknown.

However, regardless of whether the DNA anchor is a permanent feature at the centromere, the proper assembly of the kinetochore must be tightly regulated given the specific timing and location of its function. In both yeast and humans, the successful initiation of kinetochore assembly is dependent on Hsp90 (Davies *et al.*, 2010; Stemmann *et al.*, 2000). *In vitro*, Hsp90 is required for the activation of the Skp1:Ctf13 heterodimer of CBF3 (Stemmann *et al.*, 2000). Depletion of Hsp90 from reticulocyte lysates results in the inhibition of CBF3 formation around centromeric DNA in a manner dependent on the presence of nucleotide and abolished by pre-incubation with the non-hydrolysable ATP-analog AMPPNP, suggesting that a specific conformation of Hsp90 (detailed later) is required for forming competent CBF3 complexes. Furthermore, *in vitro* translated Skp1 was able to co-immunoprecipitate with Hsp90 from a reticulocyte lysate, further implicating it in the construction of CBF3 (Stemmann *et al.*, 2000). It was consequently postulated that Hsp90 requires a second chaperone, given the relative abundance of Hsp90 and the activation of Skp1:Ctf13 in the lysate. GST- and His-tagged Skp1 and Ctf13, respectively, were shown independently to have a weak interaction *in vitro*. In the presence of Hsp90 and its co-chaperone Sgt1 ('suppressor of G2 allele of Skp1'), this interaction was increased approximately 3.2-fold, suggesting the involvement of Sgt1 in this process (Bansal *et al.*, 2004). Cells expressing an Sgt1 mutant carrying point mutations in the N-terminal region were found to be incapable of forming CBF3 complexes (Kitagawa *et al.*, 1999).

Understanding the structure of Hsp90 and how this influences its function is therefore essential to deciphering the coordinated role of the Hsp90:Sgt1 chaperone machine.

1.3.1 COMPOSITION OF THE INNER YEAST KINETOCHORE

The assembly of the kinetochore is dependent on the hierarchical formation of sub-complexes at the centromere as facilitated by the Hsp90:Sgt1 chaperone machine. The point centromere of *S. cerevisiae* is composed of three conserved nuclease-resistant DNA elements termed CDEI, CDEII and CDEIII (Funk M *et al.,* 1989). These three elements combine to accommodate the DNA-binding anchors of the kinetochore: CBF3, a heteromeric complex of six subunits, a dimer of CBF1 (centromere binding factor 1) and Mif2p, with the former and latter binding to the conserved 125 bp CDEIII element (Cohen *et al.*, 2008; Westermann *et al.*, 2007). The heterochromatic centromeric region is associated with a specialised nucleosome that includes an H3 histone-like subunit (CENP-A in humans, Cse4p in yeast) in place of the canonical H3 histone (Smith, 2002) (Figure 1.17, dashed box). Centromere defect mutants of both *cse4* and histone H4 can be complemented through overexpression of the other, which suggests that the two proteins are dependent on each other to form a functional nucleosome particle (Glowczewski *et al.*, 2000).

It has been proposed that specialised CENP-A nucleosomes act as epigenetic markers for the propagation of the centromere in replicated chromosomes (Black *et al.*, 2004; Westermann *et al.*, 2007). In light of this, a recent crystal structure of the CENP-A-containing nucleosome shows that it forms a structure similar in its overall topology to the H3 nucleosome, but that, in contrast to the canonical particle, forms less stable contacts with DNA (Tachiwana *et al.*, 2011). One of the key differences between yeast Cse4p and the canonical H3 histone lies within the structure of the N-terminus, which facilitates an interaction of Cse4p with components of the linker layer complex Ctf19 (Chen *et al.*, 2000). The centromeric nucleosome therefore provides a physical link between the DNA binding components and the motility mechanisms present at the kinetochore-

microtubule interface. This N-terminal region, however, was shown to be non-essential for the localisation of Cse4p to the centromere, which is instead specified by the pre-assembly of the CBF3 complex on the CDEIII element of the point centromere (Ortiz *et al.*, 1999). The CBF3 complex therefore provides the epigenetic specification of kinetochore assembly via a specialised nucleosome particle at the centromere.

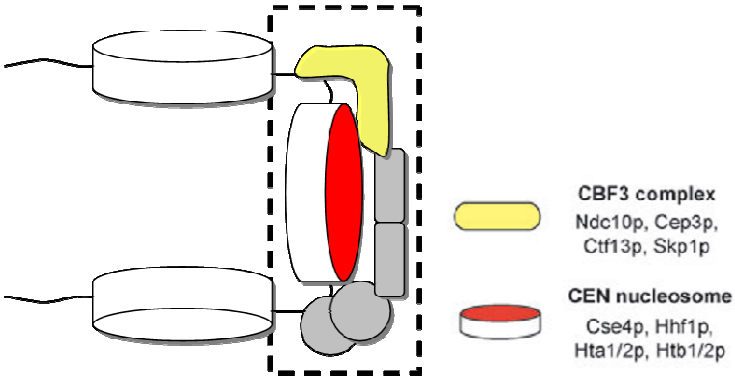


Figure 1.18: **Cartoon Schematic of the Yeast Centromere** – The primary DNA anchor, CBF3, is shown in **Light Yellow**, with the specialised centromeric nucleosome highlighted in **Red**. The structural basis for this interaction is important in understanding how the assembly of the kinetochore is initiated.

The fact that Cse4p-containing nucleosomes are diminished at sequences juxtaposed 200 bp from the centromere, emphasises the specialisation of this region of the chromosome for the propagation of kinetochore assembly (Meluh *et al.*, 1998). The structural basis for the interaction between the Cse4p nucleosome and CBF3 remains uncharacterised in yeast, but evidence from studies of the equivalent complex in humans, CCAN, suggests that the interaction occurs with the canonical H3 histone, and not with CENP-A/Cse4p (Hori *et al.*, 2008; McAinsh *et al.*, 2011).

The CBF3 complex is, therefore, one of the most important spatial determinants for kinetochore assembly at the time of mitosis, with the elements for kinetochore localisation specified structurally by the constituents of the CBF3 hexamer (Figure 1.18, dashed box).

1.3.1.1 Insights into the Structure of Ndc10

Two recent structures of the DNA binding modules of Ndc10, one a 2.8 Å structure of the *Kluveromyces lactis* homologue and the other a 1.9 Å structure of

its *S. cerevisiae* counterpart, emphasise the purpose of CBF3 as a scaffold and in facilitating the formation of the Cse4p-containing nucleosome (Cho *et al.*, 2011; Perriches *et al.*, 2012).

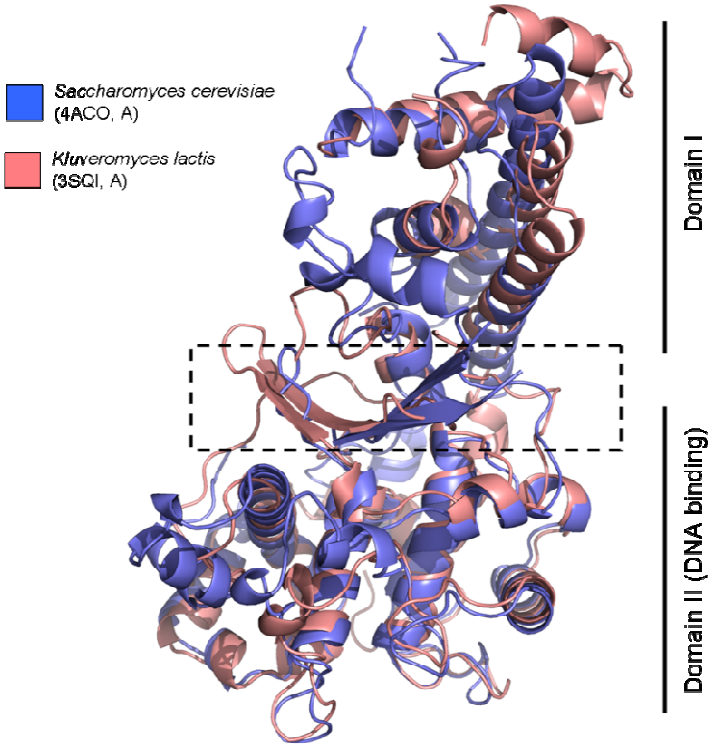


Figure 1.19: **Structural Superposition of Yeast Ndc10** (PDB IDs: 4ACO and 3SQI) – Superposition of the structures of Ndc10 homologues from *S. cerevisiae* (**Blue**) and *K. lactis* (**Pink**) shows a conservation of the overall fold in both homologues. The key differences lie in the extra loop regions of Ndc10 from *S. cerevisiae*. The crystallised fragment shows the orientation of **Domain I** and **Domain II** separated by a beta-sandwich region (**Dashed Box**). Domain II mediates DNA binding. Ndc10 was found not to have a preference for a specific sequence of DNA, suggesting that, rather than being a spatial determinant of CBF3, it reinforces its DNA binding activity together with Cep3 at CEN DNA.

Both structures show the bi-lobed structure of a proteolytically-resistant fragment of full length Ndc10, with Domain I at the extreme N-terminus (Figure 1.19, top) forming a five-helix bundle that is flanked at its C-terminus by a small beta-sandwich region (Figure 1.19, dashed box), which bridges with Domain II, the site of DNA binding (residues 101-402 in *K. lactis*) (Figure 1.19, bottom). The overall structure of the Ndc10 N-terminus is conserved between both yeast homologues (Figure 1.19). While Ndc10 has been shown to form dimers (Jiang *et al.*, 1993), neither of these structures elucidates the mode of dimerisation, though both authors speculate that this capacity to self-associate is essential to some aspect of its function at the centromere. A superposition further shows structural similarities of Ndc10 to Cre recombinase of the IB topoisomerase/λ-integrase (IB/Int) superfamily, which possesses DNA binding and catalytic activities relevant to DNA repair (Perriches *et al.*, 2012). This suggests a mode of DNA binding conserved between Ndc10 (from both *S. cerevisiae* and *K. lactis*) and

the class of IB/Int DNA-binding proteins (Figure 1.20), though the former clearly fulfils a non-catalytic purpose in the context of kinetochore assembly.

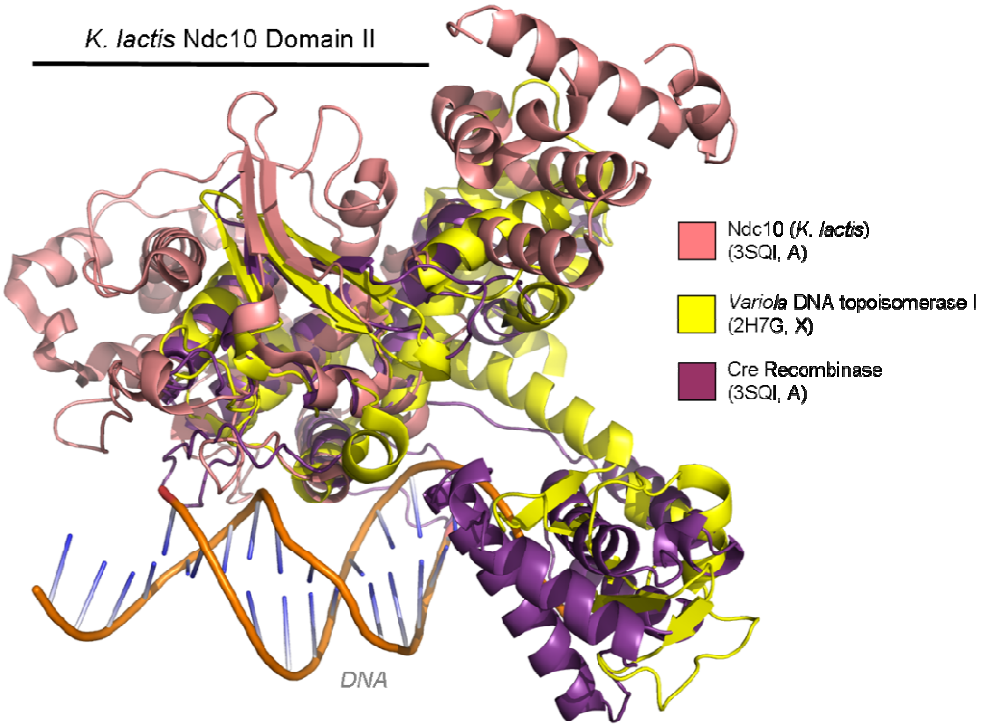


Figure 1.20: **Structural Superposition of Ndc10 and Members of the IB/Int Superfamily –** Ndc10 shares some structural homology with members of the IB/Int superfamily of recombinases and DNA-interacting proteins, in particular in the beta-sheet region adjacent to the the DNA binding domain.

Both groups concluded that Ndc10 fulfils a general scaffolding role, with no preference for sequences found at the centromere. Instead, it is directed to the site of kinetochore formation through contacts with both Cbf1 at CDEI and the CBF3 core of Skp1:Ctf13 and the Cep3 homodimer at CDEIII, with both halves of the Ndc10 dimer individually contacting Cbf1 and the CBF3 complex (Cho *et al.*, 2011). Pulldown experiments additionally suggest that Ndc10 recruits the histone-chaperone Scm3 to the nucleosome that is coiled around CDEII, which would instigate the exchange of the H3 histone with Cse4p. As such, Ndc10 is vital to the propagation of kinetochore assembly and acts as the central scaffold linking all factors of the centromere together (Cho U *et al.*, 2011). Its role in bridging the CDEI and CDEIII element appears to stimulate the assembly of additional layers of the kinetochore, rather than instigating CBF3 assembly. This role is instead reserved for its DNA-binding counterpart within CBF3, Cep3.

1.3.1.2 INSIGHTS INTO THE STRUCTURE OF CEP3

Mutants of Cep3 were first identified as temperature-sensitive isolates exhibiting a G2-M delay phenotype, implying the failure of the mitotic machinery to attach to sister chromatids and initiate chromosome segregation (Strunnikov *et al.*, 1995). A subsequent sequence analysis by the same group showed that Cep3 contained the only recognisable DNA binding motif amongst kinetochore proteins. The use of an antibody specific to Cep3 inhibits its interaction with *CEN* DNA and consequently the formation of CBF3 complexes *in vitro*, presumably by blocking the association of Cep3 with DNA and preventing CBF3 initiation (Lechner, 1994).

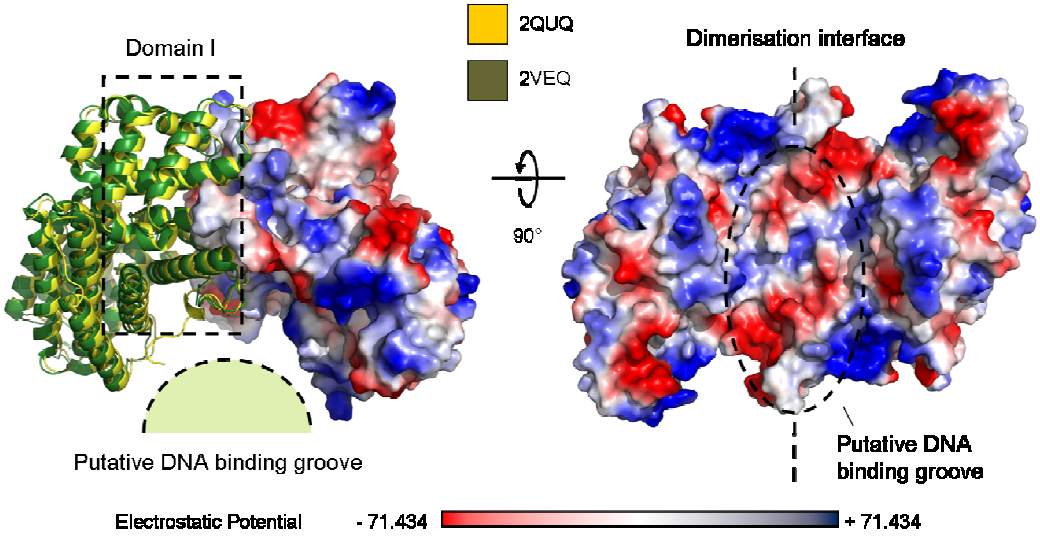


Figure 1.21: **Crystal Structure of Cep3** (PDB IDs: 2QUQ and 2VEQ) – Cep3 is the key determinant in targeting the CBF3 complex to the centromere. Two independent studies of yeast Cep3 have yielded atomic resolution structures (2QUQ, **Yellow**; 2VEQ, **Green**) of a Cep3 dimer. The dimerisation interface predominantly involves **Domain I** of Cep3. In one proposed model for DNA binding, DNA binds across the underside of the dimerisation interface, with contacts coming from both protomers in the dimer.

Cep3 belongs to a class of Zn_2Cys_6 zinc-finger domain-containing proteins, a feature that mediates its DNA binding activity on a conserved CCG site found within the CDEIII element at the centromere (Espelin *et al.*, 1997). This sequence is a common recognition site for this class of proteins and is often palindromic, appearing in a triplet configuration that facilitates interactions with the major groove of DNA (Wu *et al.*, 1995). In the CDEIII element, a single conserved CCG half-site was found to be essential for chromosome segregation, most likely due

to its role in recruiting Cep3 to the centromere and initiating kinetochore assembly, with additional contributions made by a conserved TGT sequence removed approximately 12 bp 3' from CDEII (Jehn *et al.*, 1991). While Cep3 can interact with DNA in the absence of other CBF3 components (Purvis *et al.*, 2007), a stable half-life of Cep3 on centromeric DNA depends on structural contributions from the Skp1:Ctf13 heterodimer (Russell *et al.*, 1999). Deletion of the N-terminal zinc-finger results in a loss of stable binding to CDEIII (Purvis *et al.*, 2007).

Two structures of Cep3 from *S. cerevisiae* illustrate the arrangement of the regions C-terminal to the N-terminal zinc-finger domain, which is believed to be separated from the C-terminus by a flexible linker that renders the full length protein unable to form crystals (Purvis *et al.*, 2007; Bellizzi *et al.*, 2007). The crystallised fragments (PDB IDs: 2VEQ and 2QUQ) illustrate the dimerisation interface of Cep3 and sites putatively mediating interactions with other CBF3 components. The dimer forms a symmetrical bi-lobal crescent-shape (Figure 1.21), with the interface involving extensive contacts between Domain I of adjacent protomers, in contrast to the coiled-coil mediated dimerisation interface of other DNA binding zinc-finger proteins (Marmorstein *et al.*, 1992). The channel formed by this interface, which on its underside contains patches of positive electrostatic potential (Figure 1.21), is sufficient to accommodate B-form DNA (Purvis *et al.*, 2007). The flexible linker between the dimerisation domain and zinc-finger contributes to the envelopment of the DNA duplex, allowing the zinc-clusters to contact the CCG and TGT half-sites by adopting a conformation pointing towards the centre of the pseudo-dyad (i.e. the TGT and CCG sequences) (Purvis *et al.*, 2007) (Figure 1.22). The conformation of the dimer in this model would therefore act as the distance constraint specifying binding to the CDEIII element.

By enveloping DNA in this manner, the binding of Ctf13 to Cep3 can be accommodated on the underside of the DNA duplex. Previous biochemical data has shown that Ctf13 binds halfway between the aforementioned CCG and TGT sites (Espelin *et al.*, 1997), an interaction compatible with the model of DNA

binding proposed by Purvis *et al.* (Figure 1.22). Herewith, Cep3 and Ctf13 form a stable quaternary complex, which fully engulfs the DNA, creating a stable entity with a prolonged half-life. This structural role for Cep3 offers an element of spatial control to the assembly of CBF3 components at the centromere, suggesting that, firstly, the geometry of DNA elements at the centromere permits the initiation of kinetochore assembly via Cep3, subsequent to which the initiation signal is propagated through the recruitment of an Ndc10 dimer and reorganisation of centromeric DNA to accommodate a CENP-A/Cse4p:H4-containing histone. It is possible that this spatial control is complemented by the temporal control offered by mitotic kinases, which are known to be associated with both Ndc10 and other kinetochore components (Westermann *et al.*, 2007).

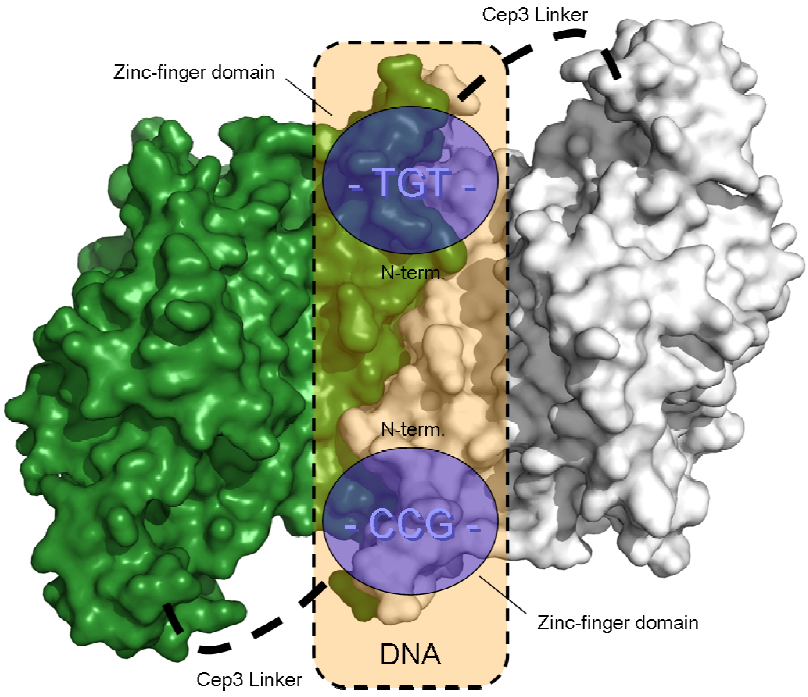


Figure 1.22: **DNA Binding Model of *S. cerevisiae* Cep3** (PDB ID: 2VEQ) – In the study conducted by Purvis *et al.*, the mode of DNA binding illustrated here was proposed to involve contacts from both protomers of a Cep3 dimer, with DNA binding within a groove formed on the underside of the dimerisation interface. This model accommodates previous observations related to the location of the N-terminal zinc-finger domain of Cep3, which were found to be important for binding the critical TGT and CCG half-sites at the centromere. The linker between the zinc-finger and C-terminal domains confers the necessary flexibility of the N-terminus to bind to DNA in this manner, with the other domains (including Domain I, the dimerisation domain) stabilising the DNA on the opposite side.

1.3.1.3 Structural and Functional Predictions for Ctf13

Little is known about the biochemistry, hydrodynamic behaviour or structure of Ctf13. Despite being critical to the structural stability of the CBF3 complex, it is highly unstable, readily degraded and therefore lacks extensive characterisation (Russell *et al.*, 1999). Phenotypic analysis of the *ctf13-30* mutant shows temperature-sensitive kinetochore defects with characteristic features of a G2/M arrest (Doheny *et al.*, 1993) that can be suppressed through overexpression of Skp1, its CBF3-specific binding partner (Hieter *et al.*, 1996). Ctf13 was shown to form a stable complex with Skp1 and a homodimer of Cep3 in size exclusion chromatography experiments, and this complex can furthermore associate with Ndc10 in the absence of CEN DNA *in vitro* (Russell *et al.*, 1999).

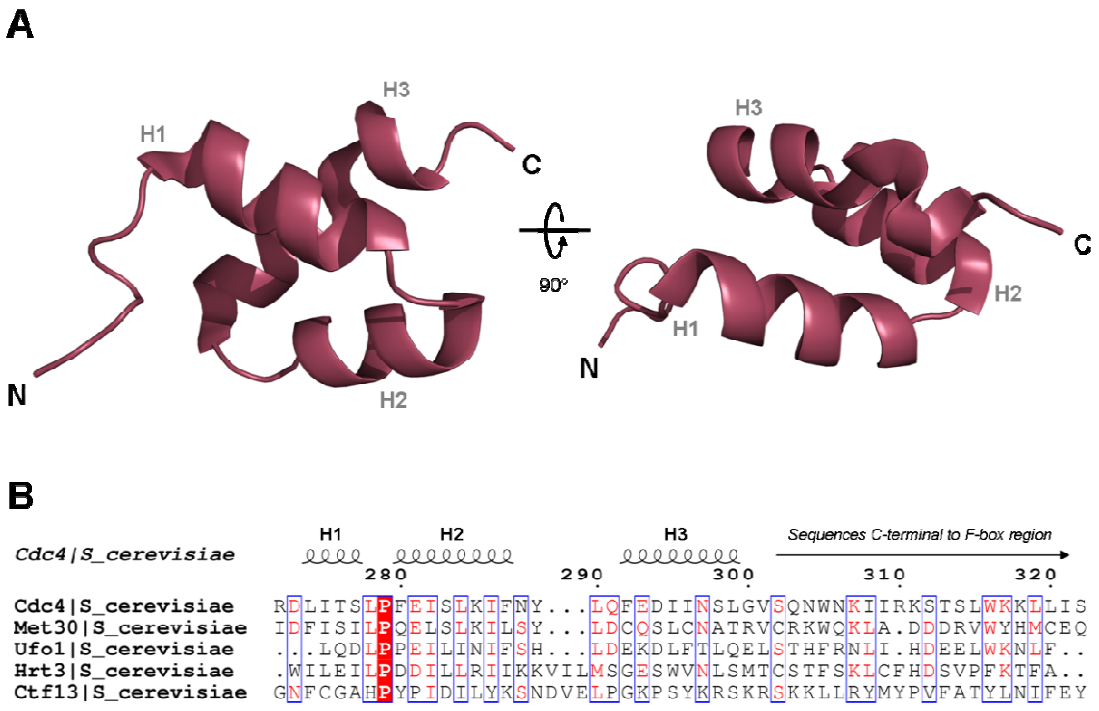


Figure 1.23: **Structure of the F-box motif of Skp2** (PDB ID: 1LDK) – **(A)** The F-box motif is a simple three helix bundle (**H1, H2 and H3**) that is recognised by Skp1 in SCF complexes and as part of the structural core of the CBF3 complex. **(B)** Sequence alignment of the putative F-box of Ctf13, with other canonical F-boxes from the yeast *S. cerevisiae*. This alignment was generated using the BLOSUM62 scoring matrix and the figure generated using ESPript (Gouet *et al.*, 1999).

Selective deletions of Ctf13 showed that an internal fragment of the protein mediates binding to Cep3 (residues 139 – 336), while the Ndc10 binding site lies N-terminal to this sequence (Russell *et al.*, 1999). This information suggests that

Ctf13 performs a scaffolding function in the absence of CEN DNA, bridging the two homodimeric DNA binding subunits of CBF3 together with Skp1. Whether or not this occurs before the initiation of kinetochore assembly, is unclear, but it is possible that Ctf13, which is a client of Hsp90 (Stemmann *et al.*, 2002), facilitates DNA binding of CBF3 after being activated by the molecular chaperone machinery.

In the absence of an atomic resolution model of Ctf13, biochemical data and sequence analyses have shed light on its domain composition. The N-terminus comprises sequences that match poorly to the consensus residues for an F-box motif. The F-box is a helical bundle (Figure 1.23) and N-terminal hallmark of molecular scavengers acting as part of the SCF ubiquitin ligase machinery to gather targets for degradation (reviewed in Jonkers *et al.*, 2009). Altering residues in Ctf13 that are weakly conserved in other F-box proteins abolishes binding of Ctf13 to Skp1 (Russell *et al.*, 1999), which itself also acts as the receptor for F-box containing proteins in the context of the ubiquitin ligase machinery (Yoshida *et al.*, 2011). This interaction is predominantly hydrophobic and stimulates the phosphorylation of Ctf13, a role believed to regulate the assembly of the kinetochore (Kaplan *et al.*, 1997). However, sequences C-terminal to the N-terminal F-box remain critical to the interaction with Skp1, as the putative F-box itself is insufficient to stimulate binding (Russell *et al.*, 1999). Further studies revealed that the F-box region is sufficient for decreasing the half-life of GST, and that a GST-fusion of full length Ctf13 increases its half-life significantly, suggesting that the availability of the F-box region for downstream functions destabilises the protein *in vivo* via the ubiquitin-ligase pathway. Both Skp1 and Cep3 are necessary to confer stability to the complex (Russell *et al.*, 1999), reinforcing the interdependence of CBF3 components for their role at the centromere.

Proteins of the F-box family are commonly associated with a C-terminal leucine rich repeat (LRR) or a beta-propeller-like WD40-repeat motif, which bind targets specified for degradation (Kipreos *et al.*, 2000). According to the CATH topology classification (CATH Code: 3.80.10.10), the region between amino acids 307-445

of Ctf13 forms an LRR motif with poor sequence identity to other LRR-containing proteins (Cherry *et al.*, 2012; Cuff *et al.*, 2011).

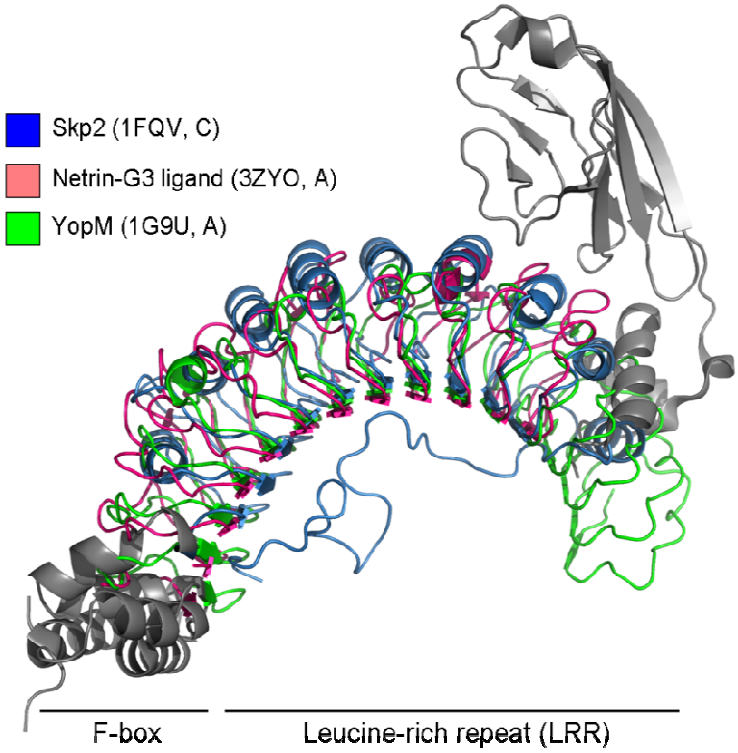


Figure 1.24: **Structure of Leucine Rich Repeat Proteins** (PDB IDs: 1FQV, 3ZYO and 1G9U) – The leucine rich repeat motif is commonly associated with F-box proteins. The core of the motif contains a helix-strand repeat pattern, with sufficient repeats resulting in a distortion of the fold into a horseshoe-like shape. The underside of the horseshoe is typically the sight for quaternary interactions. Leucine rich repeat proteins were found to be clients of the Hsp70 chaperone due in part to their hydrophobic nature. The **Grey** regions in this diagram represent sequences that are not part of the LRR motif.

An LRR has the consensus sequence 'LxxLxLxx', numerous copies of which typically result in a helix-strand repeat motif interspersed by linker regions (Figure 1.24). The number of LRR repeats and the length of the loop regions determines the overall curvature of the domain, with the concave face of the 'horseshoe' lined by beta sheets and the convex side by alpha helices (Kobe *et al.*, 2001). F-box-coupled LRRs are commonly associated with protein turnover, though the role of Ctf13 in kinetochore assembly represents an example outside of this context (Hermand, 2006). Instead, Ctf13 is targeted by Cdc4, which is itself an F-box protein involved in modulating the G1/S and G2/M cell cycle transitions, by scavenging cell cycle proteins for degradation in conjunction with Skp1 (Kaplan *et al.,* 1997; Jonkers *et al.*, 2009).

Selective deletion of sequences C-terminal to the F-box confirmed that Cep3 and Ndc10 bind to the region corresponding to the putative LRR (Russell *et al.*, 1999). A two-hybrid approach has been used to show that deletion of 58 N-

terminal residues of Cep3 from *Candida glabrata*, which reside in the zinc-finger domain, or deletion of residues 470-611 at the C-terminus, abolishes the interaction with Ctf13 (Stoyan *et al.*, 2004). The interaction between Ctf13 and Cep3 therefore involves both ends of Cep3 and an internal region of the LRR component of Ctf13. If this involves both Cep3 monomers, then the interaction site coincides with the model proposed by Purvis *et al.*, showing how the LRR of Ctf13 might straddle two Cep3 molecules, forming a toroidal quaternary complex at the centromere that stabilises the CBF3 anchor.

1.3.1.4 STRUCTURAL AND FUNCTIONAL STUDIES OF SKP1

The best characterised function of Skp1 is its role as the acceptor protein for F-boxes in the aforementioned ubiquitin ligase machinery. The purpose of ubiquitination is to mark proteins for degradation, in the event of co-translational misfolding (Turner *et al.*, 2000) or for the selective regulation of cellular processes, such as transcription (reviewed in Geng *et al*., 2012) and the recycling of membrane proteins (reviewed in MacGurn *et al.*, 2012). The general scheme of ubiquitination is carried out by three classes of enzymes, E1, E2 and E3. The E1 enzyme performs the initial activation of the 76-residue ubiquitin molecule through its adenylation and subsequent thioesterification to an active site cysteine (Lee I *et al.*, 2008; Olsen *et al.*, 2009). The activated ubiquitin molecule then undertakes a further transfer to one of many E2 enzymes, each of which interacts transiently with a selection of E3 complexes, such as the anaphase-promoting complex (APC) (reviewed in Barford, 2011), that carry the substrate being targeted for degradation (Wenzel *et al.*, 2010). Skp1 is one of three components common to the SCF (Skp1-Cullin-F-box) category of E3 ligase complexes together with Cul1 and ROC1/Rbx1, and a variable F-box protein (Figure 1.25). The purpose of Skp1 in these complexes is to stabilise the F-box and potentiate the substrate binding activity of the C-terminal LRR or WD40 protein recognition motif (Yoshida *et al.*, 2011), and ensure that the ubiquitin-moeity is transferred to the degradation target (Wenzel *et al.*, 2010). Mutations in the C-terminal F-box-interacting helices of Skp1 cause a G1 arrest at non-

permissive temperatures, demonstrating the importance of this region to the regulation of the cell cycle via the protein turnover pathway (Bai *et al.,* 1996).

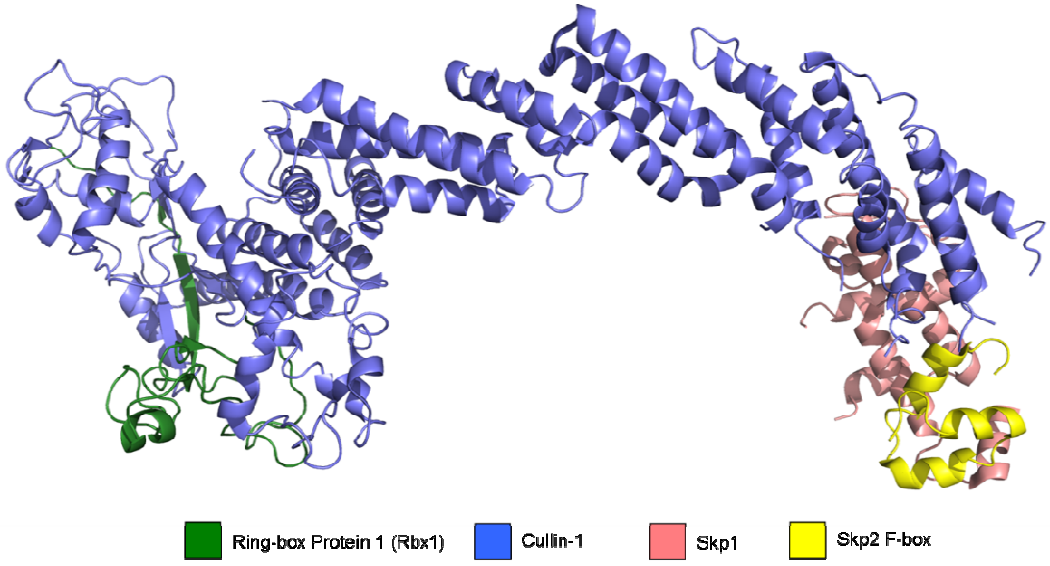


Figure 1.25: **Partial Structure of an E3 SCF Ubiquitin Ligase Complex** (PDB IDs: 1LDK and 3TDU) – The E3 SCF ubiquitin ligase complex is the selective scavenger for proteins that are being targeted for degradation. A typical SCF-type E3 ligase consists of one copy of Skp1 (**Pink**), a cullin molecule (**Blue**), a ring-box adaptor protein (**Green**) and an F-box, which, in the structure illustrated here, is from the leucine rich repeat protein Skp2 (**Yellow**). E3 ligases interact with activated E2 molecules and facilitate the conjugation of ubiquitin onto the target molecule.

The role of Skp1 in the initiation of kinetochore assembly rests in its contribution with Ctf13 to the structural core of CBF3 (Russell *et al.*, 1999). The BTB/POZ domain, comprising the largest portion of the full length protein, is a protein:protein interaction module common to many proteins functioning in multi-component cascades (reviewed in Stogios *et al.*, 2005) (Figure 1.26, A). The *skp1-12* mutant carries a mutation within the first strand (residues 4-10) of the BTB/POZ (Figure 1.26, B), leading to G2 cell cycle arrest at permissive temperatures and a chromosome missegregation phenotype as seen in other kinetochore mutants (Bai *et al.*, 1996). At semi-permissive temperatures, the same mutant displays characteristics of a G1 arrest phenotype, with a marked decrease in neddylated ubiquitin-like molecules and deactivation of the E3 ligase machinery (Siergiejuk *et al.*, 2009).

The leucine to glycine (L8G) mutation of *skp1-12* therefore destabilises the core BTB/POZ fold, with restrictive/semi-permissive temperatures encouraging the

unfolding of the three-strand N-terminus, in turn leading to a complete loss of function and G1 cell cycle arrest. Permissive temperatures, on the other hand, allow Skp1 to maintain its G1 function in regulating protein degradation, but the localised destabilisation is sufficient to prevent ternary interactions conducive to its G2 function. This suggests that the BTB/POZ domain mediates interactions relevant to kinetochore assembly *in vivo*.

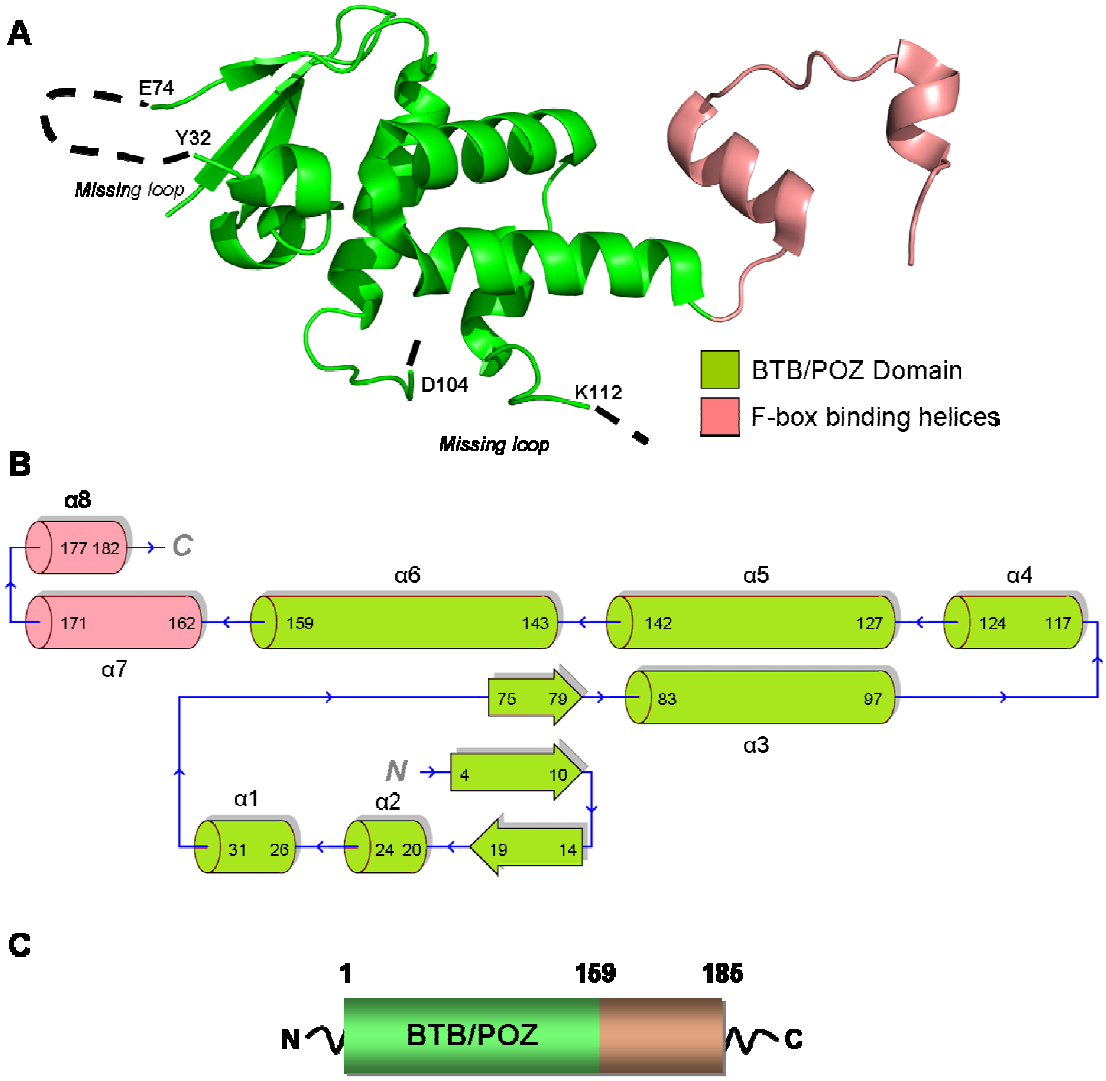


Figure 1.26: **Structure of Skp1** (PDB ID: 3MKS) – **(A)** Crystal structure of full length yeast Skp1 from the complex structure with Cdc4, an F-box containing protein. The structure (PDB ID: 3MKS) is missing the loops between residues 32-74 and 104-112, though these features are commonly missing in structures of Skp1 (see, for example, PDB entries 1LDK and 2OVP). The BTB/POZ domain makes up the largest proportion of the full length protein (**Green**) and consists of an N-terminal half with a three-strand region, and a helical C-terminal lobe. The F-box interacting helices (**Salmon**) contact F-box motifs. **(B)** Domain diagram of yeast Skp1, showing the residue range of secondary structure elements in the tertiary fold. Once again the BTB/POZ is shown in **Green**, with the F-box binding helices in **Salmon** colour. This diagram was modified from the output of PDBsum. **(C)** Linearised domain diagram diagram of Skp1.

Human Cul1 (a component of SCF complexes) is able to compete with Sgt1 for binding to Skp1, with the latter binding at a K_D of 1.64 μM and the former 24 nM (about a 70-fold greater affinity) (Zhang *et al.*, 2008). This observation suggests that Sgt1 and Cul1 share a binding site, and that their interaction with Skp1 is mutually exclusive. The structure of a human SCF complex containing human-Skp1 and human-Cul1 (PDB ID: 1LDK) shows that this interaction is mediated by the BTB/POZ domain alone, and that binding occurs across the N-terminal lobe, with contributions from residues in α5 and the loop region proceeding α3 (Zheng *et al.*, 2002) (Figure 1.27). Therefore, while the F-box binding helices are relevant to kinetochore assembly through the F-box of Ctf13, the BTB/POZ domain also entails vital functions for averting G2 arrest.

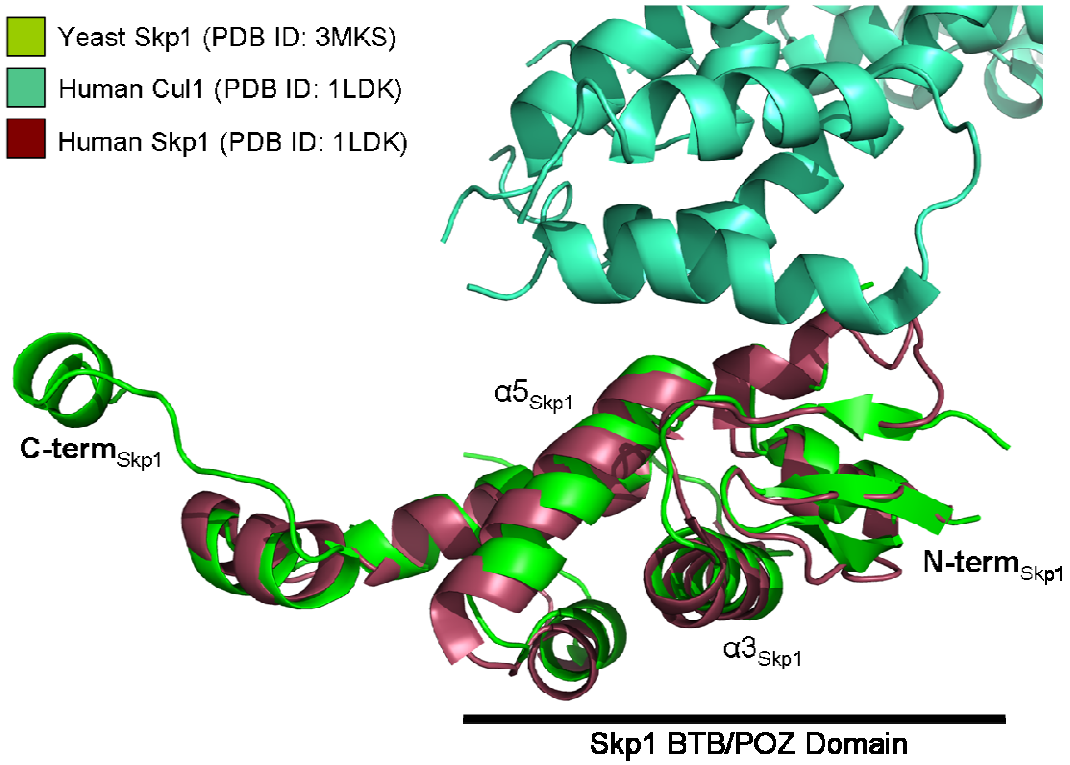


Figure 1.27: **Structure of the SCF Components Cul1:Skp1** (PDB ID: 1LDK) – The crystal structure of human Skp1 in complex with Cul1 (**Green Cyan**) shows an example of an interaction mediated by the BTB/POZ domain of Skp1 (yeast Skp1 in **Green**, human Skp1 in **Dark Red**). Contributions from α5 and the loop region after α3 of Skp1 are sufficient to outcompete Sgt1 for binding to Skp1. The existing biochemical data, together with this structure, places the interaction of Sgt1 within this region.

The structure of Skp1 in complex with Cdc4 (PDB ID: 3MKS) shows the mode of interaction of the C-terminal F-box binding helices with an F-box containing protein, akin to the interaction expected between Skp1 and Ctf13 (Orlicky *et al.*,

2010). The interface between these two proteins is a hydrophobic patch that forms from helical segments of either protein interspersing with those in its binding partner, such that the F-box motif (Figure 1.23) is cradled by the C-terminal helices of Skp1 (Figure 1.26). Helices α6, α7 and α8 of Skp1 (Figure 1.26, B) pack against helices α1, α2 and α3 of the Cdc4 F-Box, an interaction that might be mimicked in the Skp1:Ctf13 heterodimer. Residues in the core of Skp1 (including Gly129, Arg136 and Ile141 in *S. cerevisiae*) are highly conserved, and interact with equally conserved residues in the F-Box of Cdc4 suggesting that these interactions are similar in other Skp1:F-Box complexes. However, as F-boxes are partnered with other protein:protein interaction motifs (Bai *et al.*, 1996), the relative affinity of Skp1 for other F-box proteins will differ depending on the involvement of these motifs in stabilising the F-box itself or Skp1. In the structure of Skp1 with the F-box protein Fbw7 (Hao *et al*., 2007), sequences immediately prior to its WD40 motif stabilise the F-box (PDB ID: 2OVP), while the C-terminal tail end of the LRR of Skp2 (Hao *et al.*, 2005) turns back across the concave face of the horseshoe to interact with the N-terminal F-box (PDB ID: 2AST). Both of these structures serve as examples of how sequences C-terminal to the F-box confer greater stability to the N-terminus and thereby support the interaction with Skp1.

Due to its role in both protein turnover and kinetochore assembly, Skp1 could provide a plausible link between activation and degradation of the CBF3 complex. Phosphorylation of Ctf13 leads to its activation and the formation of CBF3 complexes *in vitro*, stabilising an otherwise short half-life (Kaplan *et al.*, 1997). This modification can be reversed by phosphatase treatment following CBF3 assembly *in vitro* (Kaplan *et al.,* 1997; Rodrigo-Brenni *et al.*, 2004), which highlights the reversible nature of these complexes and their partial dependence on the phosphorylation state of Ctf13. Rapid degradation of Ctf13 is thought to occur via Cdc4, an F-box containing protein that functions in the ubiquitin ligase machinery (Kaplan *et al*., 1997), providing the link to the protein turnover pathway in the scheme of factors regulating the kinetochore.

Critical to the rapid turnover of CBF3 is an interaction between Skp1 and the Hsp90 co-chaperone Sgt1 (Rodrigo-Brenni *et al.*, 2004). As Ctf13 is a client of Hsp90 (Stemmann *et al.*, 2002), the initiation of kinetochore assembly depends on the post-translational modification of CBF3 components as stimulated by Skp1 and the molecular chaperone machinery. The putative model for the assembly of the kinetochore therefore involves the stabilisation of a Skp1:Ctf13 heterodimer by an Hsp90:Sgt1 chaperone pair, which can subsequently bind Ndc10 and Cep3, a step that makes the CBF3 complex competent for binding CEN DNA when coupled to the phosphorylation of its constitutents.

1.4 MACROMOLECULAR CRYSTALLOGRAPHY

X-ray crystallography is a prolifically used method that has made significant contributions to our understanding of biology through the elucidation of atomic or near-atomic resolution structures of biological macromolecules. Technological advents such as charge-coupled device (CCD) detectors (Hammersley *et al.*, 1997) and significant developments in crystallographic software packages, such as PHENIX (Adams *et al.*, 2010) and the CCP4 suite (Winn *et al.*, 2011), have led to substantial automation in the pipeline of structure determination. Key to this is the automation of data processing at synchrotron sources, such as the *XDS* (Kabsch, 2009) pipeline at DIAMOND synchrotron, which combines spot identification (*COLSPOT*), crystal parameter determination (*IDXREF*), data integration (*INTEGRATE*) and scaling (*XSCALE*) into a streamlined real-time process with data collected live on a beamline. In this way, structure solution from a dataset of reasonable quality can occur using the resources at a synchrotron source, whether by multi-isomorphous replacement (MIR), single- and multi-wavelength anomalous dispersion (SAD and MAD) experiments, or through molecular replacement (MR).

As implied by its name, X-ray crystallography fundamentally exploits the physical properties of X-rays and a crystal system. A crystal is a lattice system of repeating units (Figure 1.28). This lattice can be described in terms of its dimensions, morphology and the relationship of the repeating units to each other

and in relation to a set of parallel *planes* in space. This repeating unit is referred to as the *unit cell* and the description of its morphology and symmetry in the crystal is called the *space group*. The smallest repeating unit within the unit cell is called the *asymmetric unit*, and it comprises the unique set of objects that describe the unit cell through the symmetry operations of the space group. The two aims of protein crystallography are to provide a mathematical description of a crystal system composed of protein molecules arranged in three dimensions, and a physical description of the arrangement of atoms (from solvent, amino acids and nucleic acids). A crystal has the inherent property of containing the same information in multiples; the diffraction from a single unit cell is too weak to be measured, but the signal is amplified in a crystal. To this end, crystallographers will also exploit any inherent symmetry in a molecule to aid in their description of these components (also called *non-crystallographic symmetry*), and will limit this characterisation to the asymmetric unit.

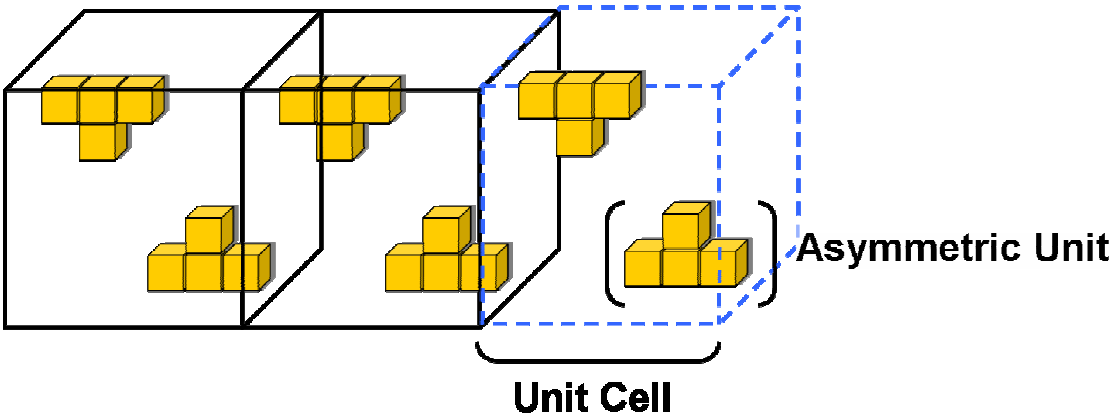


Figure 1.28: **Crystal Systems** – A crystal is a symmetrical set of repeating units, where each unit can be rotated or translated onto another by mathematical operators. The basic repeating unit is called the *unit cell*, while the smallest set of unique objects is called the *asymmetric unit*.

1.4.1 X-RAYS AS PROBES OF A MACROMOLECULAR CRYSTAL

X-rays are a form of electromagnetic radiation with a wavelength, λ, of 0.01 to 10 nanometers, which coincides with the spacings between atoms in the unit cell. When an X-ray beam (the *incident beam*) enters a crystal system it is partly absorbed by the objects in its path and either coherently or incoherently scattered, resulting in diffraction or radiation damage to the crystal. The X-ray

radiation, or *wave*, is therefore altered through its interaction with the objects in the crystal. An atom will scatter X-ray radiation and produce secondary waves in all directions, but most of the scattered waves will be subject to *destructive interference*, which results in the scattered waves being cancelled out. The conditions that lead to *constructive interference*, which reinforces the resulting wave and produces diffraction, are defined by Bragg's law of diffraction (Equation 1.1):

Eq. 1.1 $$2d \sin \theta = n\lambda$$

where θ is the incident angle of the wave and d the spacing between lattice planes on which atoms reside. This law states that only atoms that are related by an integral number of wavelengths will produce diffraction, meaning that the wavelength of X-rays, which is suited to measuring inter-atomic distances, is critical to producing constructive interference. The extent to which the wave is altered depends on the scattering propensity (the *atomic scattering factor*) of the object in its trajectory and the location of the object. The *diffraction pattern* commonly associated with X-ray crystallographic experiments is therefore a reflection of the objects in a crystal (Figure 1.29).

A single *diffraction spot* is the result of positive reinforcement of scattered waves from atoms lying relative to a set of parallel planes, which outweigh the destructive contributions of all atoms (since destructive interference does not produce diffraction).

As X-rays can be described as a wave, and therefore have a periodic property that can be expressed as a wave function, the diffraction pattern of a protein crystal contains the information needed to describe the trajectory of the X-ray through the crystal. A wave is a periodic function (Figure 1.30). In a simple one-dimensional wave for which every point on one axis, *x*, has a corresponding term on the *y* axis, *F(x)*, this means that any value of *F(x)* reoccurs with a periodicity that is specified by the function.

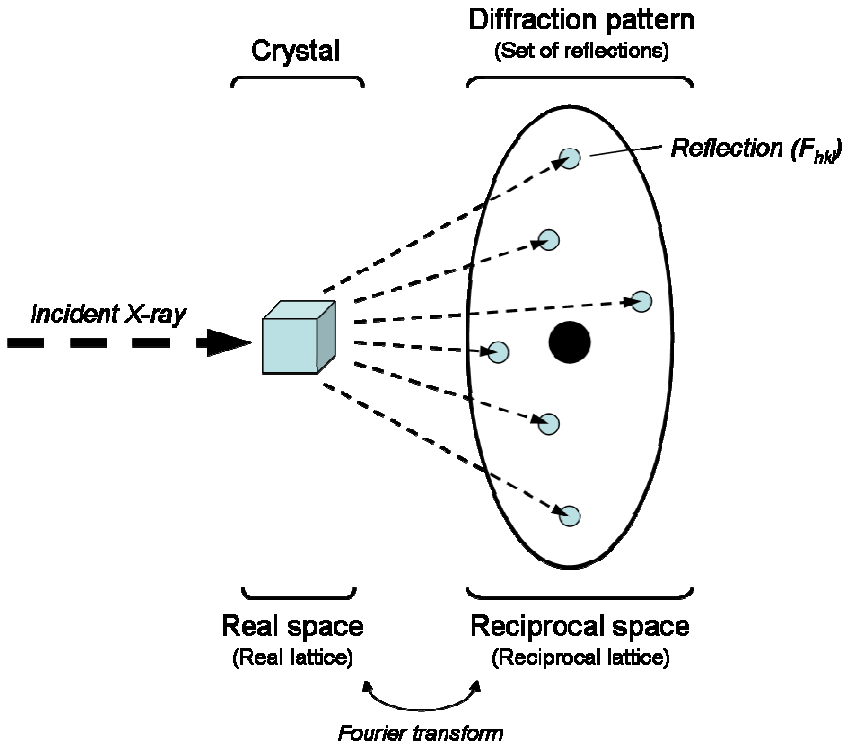


Figure 1.29: **Wave Functions as Descriptors of Crystal Systems** – A diffraction pattern of spots is produced by the passage of an incident X-ray beam through a crystal system, and its interaction with the units that make up that crystal. These spots, or reflections, are a product of the scattering power of the individual atoms that make up a crystal, with each reflection F_{hkl} corresponding to all atoms that effect the scattering along an identical set of lattice planes *hkl*. The lattice itself exists in real space, while the diffraction pattern provides a description of the crystal in reciprocal space. The large blank area typically seen at the centre of a diffraction pattern corresponds to the back stop, which is positioned to restrict damage to the detector by the direct X-ray beam.

A wave has three different properties that define the nature of this periodicity: frequency (or wavelength), amplitude and phase. The frequency defines this periodicity i.e. how frequently the terms *F(x)* are repeated, which in turn defines the length of a single repeat (the wavelength) (Figure 1.30, left). The amplitude specifies the maxima and minima of the wave over this frequency (Figure 1.30, centre), while the phase describes the position of the wave from its origin (Figure 1.30, right). The precession of the *wave fron*t around an axis traces a circle, so the phase also describes the angular position of the wave from the origin; that is, the degrees or radians that the wave has moved from the origin (Figure 1.31).

A simple two-dimensional wave can be expressed as a *cosine* function:

Eq. 1.2 $$F(x) = f_h \cos 2\pi(hx + \alpha_h)$$

Where f_h is the amplitude reoccurring at frequency h, with the wave displaced with respect to the origin of the coordinate system by phase α_h. Every term $F(x)$, is defined by the wave form described herewith.

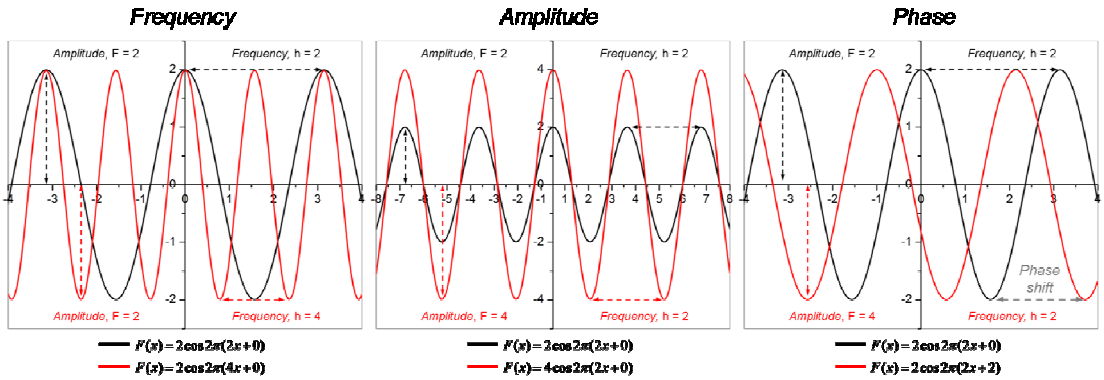


Figure 1.30: **Factors that Define the Shape of a Cosine-Wave Function** – A wave is periodic, which means that points on it or trends seen within it repeat over a period that is defined by its function. A wave has three properties that dictate this behaviour. Changes in the frequency result in changes in the periodicity of the wave, which is equivalent to the wavelength of the cosine function (**Left**). The amplitude specifies the height of the wave with respect to the perpendicular axis (**Centre**), while the phase is a factor of displacement with respect to an arbitrary origin (**Right**), and can in some ways be thought of as the movement of the wave on the axis perpendicular to the direction of the amplitude.

The contribution of a single atom to the amplitude of a constructive wave (i.e. its atomic scattering factor) is determined by the size of its electron shell, and is often expressed as $f_{j,,}$ where j is an atom type. Both the frequency and the phase are important in determining whether a wave scattered by a single atom will cause constructive or destructive interference. In a complex system like a protein crystal, the many different atom types will reinforce or weaken diffraction in a manner proportional to their scattering mass and their relative position to other atoms; in turn, every scattering factor f_j has a different numerical value specific to the atom type. The wave that exits a protein crystal to produce a single diffraction spot is a composite of waves that has been subject to the scattering power of every atom in its trajectory.

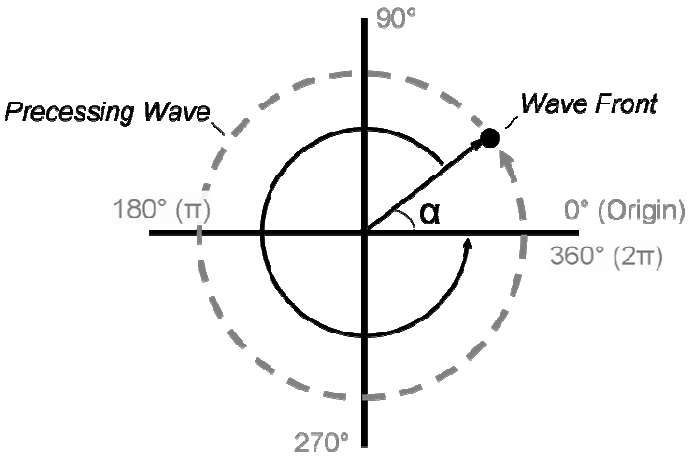


Figure 1.31: **A Wave in Radians** – A wave precesses around an axis in three dimensions, and essentially traces out a circle that is reflective of its periodicity. The phase of this wave is its displacement from the origin, and is equivalent to the angle between the wave front and the axis around which the wave is precessing.

The diffraction pattern can be represented mathematically as the sum of all the individual trajectories, as if each of these masses all existed in isolation, with the sum of those waves being equivalent to what is ultimately represented in a diffraction pattern (where each diffraction spot represents an exit trajectory for the incident wave). This function is a *Fourier sum* of all individual wave functions (as in Equation 1.1) necessary to describe a complex wave:

Eq. 1.3

$$F(x_j) = \sum_{j=1}^{n} f_j e^{2\pi i(hx)}$$

The basic function $f_j e^{2\pi i(hx)}$ is identical to Equation 1.2, but the general term for the amplitude f_h has been replaced by the scattering factor f_j and the equation simplified based on trigonometric equalities. Here, n is the number of terms addressed by the total sum, starting with $j = 1$. This Fourier sum accounts for every amplitude value f_j occurring at a position specified by the frequency h and the phase implicit in the exponential. In protein X-ray crystallography, this wave describes the effect of every atom and its electron shell on the incident beam, with each electron shell being treated roughly as a sphere. This can also mean that, for every reflection, a specific atom might have a significant effect or no effect at all on the characteristics of the wave, but must nonetheless be added to the expression to be consistent with the crystal system. Atoms that occur on a set of lattice planes will contribute more to this equation than ones that are slightly askew of that set of planes in space (for example, atoms that occur in between the planes). Since what is recorded is this summative wave, and not the

individual wave, a crystallographer must effectively work backwards to decipher the constituent waves.

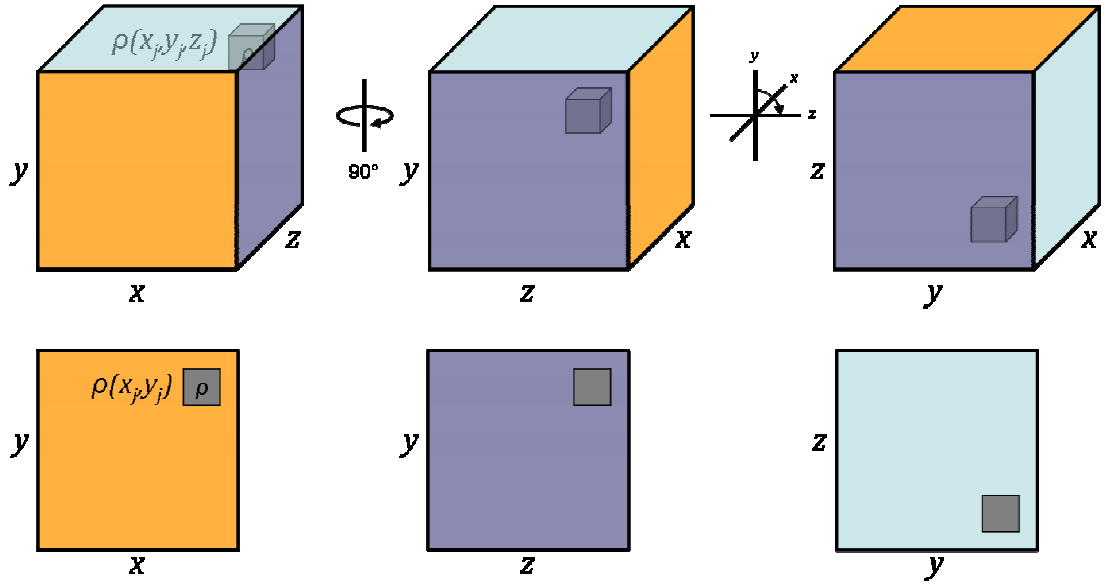


Figure 1.32: **Wave Functions as Descriptors of Crystal Systems** – A Fourier sum that accurately describes the atoms in a protein crystal must account for the wave scattering in three dimensions, as much as the coordinates of an atom in space carry three spatial coordinates. This figure highlights the necessity of a three-dimensional coordinate system, since a two-dimensional wave does not accurately describe the position of an atom in space (compare **Left** against **Centre**).

Because this function describes a two-coordinate system with one dependent variable (for every known value of *x,* there is a calculable value of *F(x)*), it is insufficient for describing a system as complex as a protein crystal (Figure 1.32). Since we want to describe something that carries a position in three spatial dimensions, a wave function of this manner must describe this type of relationship three times, such that every point along three axes, *x, y* and *z*, has a corresponding point on each of the other two axes.

The complete diffraction pattern, obtained by rotating and tilting the crystal on an axis relative to the incident beam, contains the information needed to obtain a three-dimensional coordinate system *(x,y,z)* for every centre of scattering mass *ρ*. The intensity of each diffraction spot (or reflection) also informs us about the scattering power of every centre of mass that coincides with the incident wave, with the sum of these being proportional to the square root of the measured

intensity. The Fourier sum that expresses this is the sum of each of the waves originating from each of the three coordinate systems *x*, *y* and *z* (Equation 1.4).

Eq. 1.4
$$F(x_j, y_j, z_j) = \sum_{h_j} \sum_{k_j} \sum_{l_j} f_j \, e^{2\pi i(hx_j + ky_j + lz_j)}$$

This Fourier sum describes the wave form as impacted by the atom *j* through its scattering factor f_j, and reflects the coordinates of this atom (x_j, y_j and z_j) in a fractional coordinate system in real space. The terms h_j, k_j and l_j correspond to the set of reciprocal space lattice planes. The diffraction spot (i.e. the reflection, $\rho(x_j,y_j,z_j)$ or, more commonly, F_{hkl}, referring to the reflection *F* with coordinates *hkl*) is described by the *atomic structure factor* (Equation 1.5), which is essentially equivalent to a Fourier sum with a single term for the location of the atom in reciprocal space:

Eq. 1.5
$$f_{hkl} = f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

In the case of multiple atoms that contribute to the diffraction spot, the Fourier sum has *n* terms equivalent to the number of atoms. Therefore, every reflection has a descriptor of the form:

Eq. 1.6
$$F_{hkl} = \sum_{j=1}^{n} f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

Which states that each reflection, F_{hkl}, is the sum of every atomic structure factor of every atom along a set of parallel planes. This is referred to as a *structure factor* and is specific to each reflection on the diffraction pattern. The manner in which we can describe an X-ray after it passes through a crystal therefore reinforces its use as a probe of a lattice system.

However, while these equations describe the resulting wave, they reveal nothing about the volumes that actually make up each of the objects along that wave. To decipher the contribution of an atomic volume and trace its outline, we utilise the integral of the Fourier sum illustrated in Equation 1.6 with respect to the

total volume, *V*, of the *unit cell* of the crystal. Repetition and symmetry operations within the unit cell in three spatial dimensions describe the entire crystal. The unit cell therefore contains all of the atoms that contribute to the structure factor F_{hkl}.

Eq. 1.7 $$F_{hkl} = \int_V \rho(x,y,z) e^{2\pi i(hx+ky+lz)}\, dV$$

Where the integral *dV* is the shortened expression for the integral *dx dy dz*, which corresponds to all of the volume elements at coordinates (*x, y, z*) (i.e. all of the atoms in the unit cell), the product of which is essentially the unit cell volume. This is equivalent to the sum of peak *areas* on a curve of two-dimensions, except that, by virtue of being integrated on three axes, this equation represents a volumetric outline in reciprocal space, rather than an area. This integral therefore tells us about the volumes themselves, but as an outline rather than a direct measurement.

To this point, all of the listed equations describe the set of reflections and their role as measurements in reciprocal space of objects existing in real space i.e. within the crystal lattice itself. Equation 1.7 calculates an outline of all of the atomic volumes addressed by the Fourier sum in Equation 1.6. The true volumetric description of the *electron density* (the sum of all volumes in real space) is derived from the Fourier transform of Equation 1.7 (i.e. it is the inverse Fourier transform of the structure factor equation):

Eq. 1.8 $$\rho(x,y,z) = \frac{1}{V}\sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)}$$

Where $\rho(x,y,z)$ is the Fourier transform of F_{hkl}, and vice-versa. This equation is very similar to Equation 1.6, with every variable in this Fourier sum, corresponding to an atom in the unit cell, bearing an amplitude F_{hkl}, a frequency (*h*, *k* and *l)* (i.e. the coordinates and spacing with which F_{hkl} occurs on the lattices (*h*, *k*, *l*)), and a phase, implicit in the exponential $-2\pi i(hx+ky+lz)$. This is the equation that needs to be satisfied for structure solution, and is approximated

through the iterative reconstruction of an atomic model until it adequately describes the observed data.

A wave can also be expressed as vectors, with the length of the vector equal to the amplitude ***F*** of the wave, and the angle between the vector and the *x*-axis equivalent to the phase *α* (Figure 1.33). As the amplitude is the square-root of the intensity of the reflection (the diffraction spot), the length of the vector is directly measureable from a diffraction experiment.

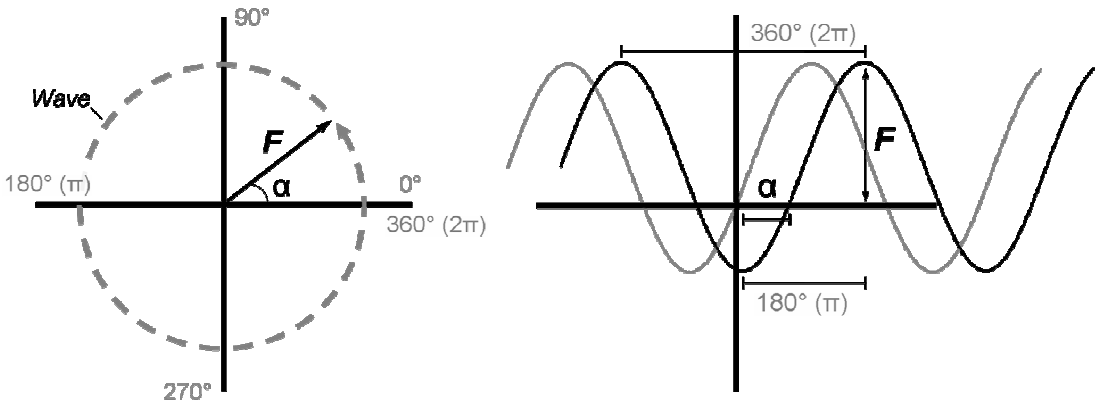


Figure 1.33: **Structure Factors as Vectors** – Partial wave functions can be describe as vectors, with the amplitude equivalent to the length of the vector (i.e. it is specifying the distance to a position on the wave from the origin) and the angle between the vector and the origin equivalent to the phase. When an X-ray arrives at the detector, it arrives in a manner that reflects its amplitude and its phase; however, the phase remains implicit in the diffraction spot because it cannot be measured directly. The goal of solving the phase problem is to determine this implicit phase.

This form of wave expression means that every atomic structure factor, which is specific to every atom that makes up a single reflection, can be summed to yield a vector description of the reflection itself (a *molecular structure factor*) (Figure 1.34), specifying the phase of atoms along an equivalent set of lattice planes in real space. Some atoms will also contribute to the molecular structure factor by lying between two sets of lattice planes, and their contribution to the amplitude of the reflection is accounted for by this function as well.

As the wave precesses around an axis, the wave vector corresponding to the leading edge of the wave precesses with it (Figure 1.33). Therefore, when the wave coincides with the detector, that vector is reflective of its phase with

respect to an arbitrary origin, specified by distances between the X-ray source and the crystal, and the crystal and the detector. Every individual reflection and the relationship of individual reflections to others can be represented by a vector, then every reflection can also be related by a vector. Expressing the atomic structure factors of volume elements in space in this way allows us to solve the *crystallographic phase problem* (see Section 1.4.3), as these relationships are related to the arrangement of atoms in a crystal.

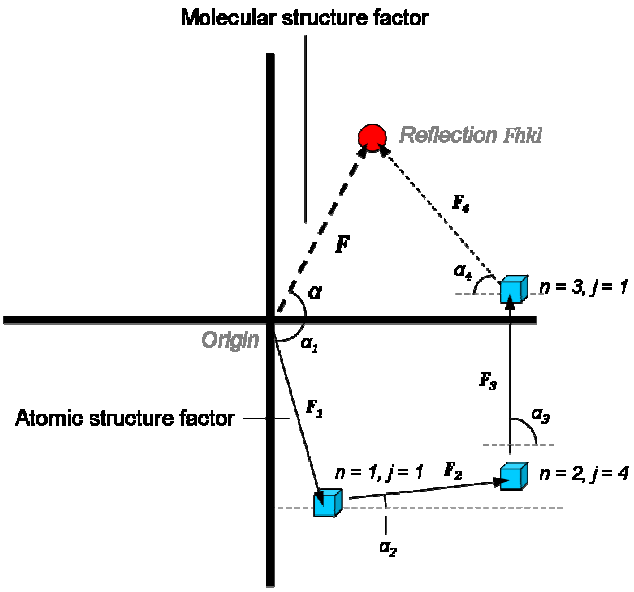


Figure 1.34: **Summing Vector Representations of Atomic Structure Factors into Molecular Structure Factors** – The sum of vectors representing individual wave functions can be expressed as a summative vector called the molecular structure in X-ray crystallographic terms. The atomic structure factor refers to a constituent vector that makes up a component of this sum. Every vector has a phase angle and amplitude, which must be determined during the phasing process.

1.4.2 Describing the Crystal Lattice

The first step in structure determination is to use the diffraction data to produce a set of *indices* that describe the sets of lattice planes *(hkl)*, which are dependent on the orientation of the crystal in the beam, and to calculate the inherent symmetry within the crystal system (its space group) and its unit cell dimensions. The space group is determined specifically by the packing of molecules within the unit cells.

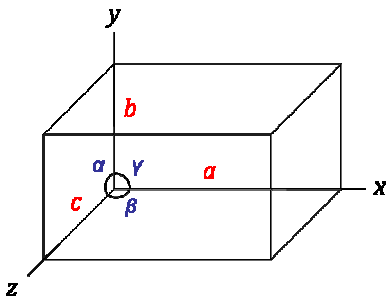


Figure 1.35: **The General Geometry of a Crystal System** – A crystal system has six different parameters that define its geometry: three angles (*α, β* and *γ*) and three sides (*a, b* and *c*).

Different crystal morphologies, which obey different geometries as defined by three faces of the crystal (*a, b* and *c,* one for each spatial dimension, *x*, *y* and *z*) and three angles (*α, β* and *γ*, one for each pair of faces) (Figure 1.35), are associated with a fixed number of space groups (Table 1.2). A certain crystal geometry, whose centres of symmetry correspond to the edges at the junction of its faces, can be described by a series of symmetry operations performed on the objects at those centres of symmetry. The lengths of the three dimensional axes, the product of which specifies the area of its faces *a*, *b* and *c*, are also referred to as *lattice vectors*.

Crystal System	**Variations**	**Unit Cell Parameters**		**Possible Space Groups**
		Faces	*Angles*	
Triclinic	Primitive	a ≠ b ≠ c	α ≠ β ≠ γ ≠ 90°	2
Monoclinic	Primitive, Base-Centred	a ≠ b ≠ c	α = γ = 90° β ≠ 90°	13
Orthorhombic	Primitive, Body-, Base- and Face-Centred	a ≠ b ≠ c	α = β = γ = 90°	59
Tetragonal	Primitive, Body-Centred	a = b ≠ c	α = β = γ = 90°	68
Trigonal (Rhombohedral)	Primitive	a = b = c	α = β = γ ≠ 90°	7
Trigonal/Hexagonal	Primitive	a = b ≠ c	α = β = 90° γ = 120°	45
Cubic	Primitive, Body-and Face-Centred	a = b = c	α = β = γ = 90°	36

Table 1.2: **The Geometric Conditions for The Seven Crystal Systems**

Of the seven crystal systems, four are also associated with additional *face (F)*-, *base- (a, b* or *c,* depending on the sides involved in symmetry) or *body (I)*-centred points of symmetry, which together form the 13 *Bravais lattices*, the set of unique symmetry descriptors needed to define the unit cell. A crystal system without these additional symmetry descriptors, where the sole centres of symmetry are

the edges of the unit cell, is said to be a *primitive* lattice (every crystal system has a primitive Bravais lattice) (Figure 1.36).

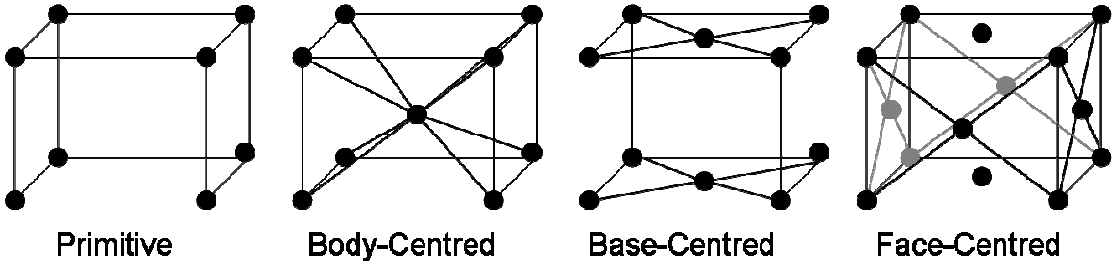


Figure 1.36: **The Set of Bravais Lattices of an Orthorhombic Crystal System** – A crystal lattice can have primitive (P) symmetry, where only the corners of the system correspond to elements of symmetry. However, the set of Bravais lattices has further body types that describe the symmetry within a crystal system, including body (I)-, base (a, b or c depending on the faces involved)- and face (F)-centred symmetry.

Each of these sets of parameters describes the repeating unit of a simple crystal geometrically. The set of mathematical operators that actually describes the symmetry is the *point group*, which is specific to the crystal system and its Bravais lattices. In total, there are 32 unique point groups, the annotation of which is specified by the conventions of the Hermann-Mauguin notation (Table 1.3) (Vainshtein, 1994).

Crystal System	**Point Groups**
Triclinic	$1, \bar{1}$ (e.g. P1)
Monoclinic	$2, m, 2/m$ (e.g. C2)
Orthorhombic	$222, mm2$ (*or* $2mm$ *and* $m2m$), mmm
Tetragonal	$4, 422, \bar{4}, 4/m, 4mm, \bar{4}2m, 4/mmm$
Trigonal (Rhombohedral)	$3, \bar{3}, 32^2, 3m^2, \bar{3}m^2$
Trigonal/Hexagonal	$6, 622, \bar{6}, 6/m, 6mm, \bar{6}2m, 6/mmm$
Cubic	$23, 2/m\bar{3}, 432, \bar{4}3m, m\bar{3}m$

Table 1.3: **List of Possible Point Groups of the Seven Crystal Systems** – The notation of these point groups is written in the Hermann-Mauguin notation, where *n* specifies the number of rotational axes through which the symmetry unit is rotated and *m* corresponding to a mirror plane, across which the symmetry unit can be mirrored. The notation *n*-bar points to rotoinversion symmetry operator.

Every number, *n*, specifies the number of *rotational axes* separated by an angle equivalent to 360°/*n* and *m* a mirror plane, that is, an imaginary plane that denotes the symmetry operation needed to superimpose symmetry-related

objects (Figure 1.37). The notation $\bar{n}$ (pronounced '*n*-bar') refers to a rotoinversion axis, which is an improper rotation operation where any given point is first rotated through *n* axes and then inverted through a mirror plane that is perpendicular to the rotation axis (Figure 1.37). All three of these operators are *symmetry elements* that can be imposed on any point in the crystal system to determine its symmetry-related partners.

In addition to these symmetry elements, space group annotations carry additional information on *screw axis* operations which are a combination of rotational and translational operations for points in the crystal system. The space group is therefore a complete designation for the symmetry and geometric elements that compose the crystal lattice.

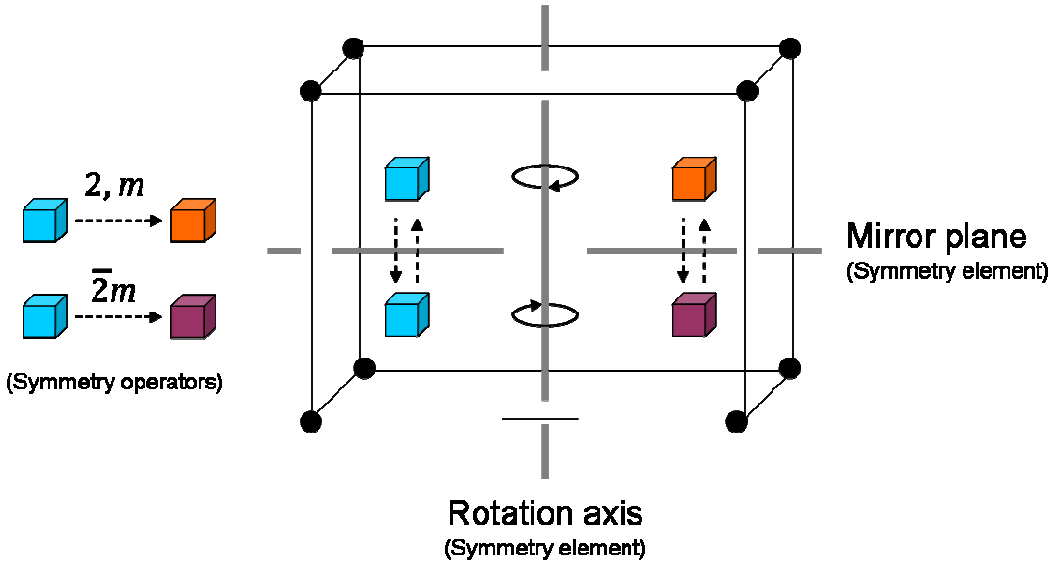


Figure 1.37: **Symmetry Elements and the Hermann-Maguin Notation for Point Groups** – The Hermann-Maguin notation is a short-hand form for symmetry operators in three dimensions. While the *n* and *m* notations, for rotational and mirror axes, are self-explanatory in their function, illustrating the rotoinversion operator illustrates its purpose as a combination of both rotational and mirror symmetry operations simultaneously (**Cyan** superimposed onto **Violet**).

Space group determination is important in generating a reliable set of reflections that describe the entire protein crystal. A set of reflections belonging to any given point group must satisfy a number of conditions in order to be classified as belonging to a particular crystal system. These conditions are specified by the symmetry operator (the point group). By their very nature, two different symmetry operators might accurately predict the occurrence of diffraction spots

if the crystal belongs to a certain space group, though lower order symmetry operators might fail to describe some percentage of the diffraction pattern. In addition, *systematic absences* in the data (that is, a pattern of reflections that are absent throughout the data but might be expected for the symmetry operator) result from the presence of screw axes in the crystal system. Software such as iMOSFLM (Battye *et al.*, 2011) and the XDS pipeline (Kabsch, 2009) are designed to systematically probe every possible set of rules pertaining to every space group. Once these are decided, the dataset is statistically evaluated based on the average intensity of the diffraction spots and the reliability of their measurement through a small angle in space. It is during this process that the data is contained within manually specified resolution limits; the larger the boundary of the diffraction pattern, the greater the resolution due to the reciprocal relationship of the data and the crystal lattice. Since the lattice spacings and unit cell dimensions are contained within this data, these programs additionally process the geometry necessary to describe the crystal system itself.

Lastly, because of radiation damage to the crystal from the incident X-ray beam, bad areas and obstructions on the detector (such as the *back stop*, designed to prevent damage to the detector by the direct X-ray beam), and areas producing poor diffraction within the crystal, the diffraction intensities must be *scaled* to correct for this error in intensity measurements, which might typically result in differences in the intensities of symmetry-related reflections. This process of *scaling,* performed in standalone programs such as *Scala* (Evans, 2006) or XDS (via the internal program *XSCALE*; Kabsch, 2009), involves firstly scaling symmetry-related reflections between diffraction images based on the consequences of diffraction and then scaling the entire dataset based on the average of all corrected intensity measurements. Once this is completed, the structure factor amplitude of all reflections (i.e. the square-root of the measured diffraction intensity) can be tabulated together with the indices of the reflections and subsequently used in addressing the *crystallographic phase problem*.

1.4.3 THE CRYSTALLOGRAPHIC PHASE PROBLEM

As the electron density equation describes a wave, the crystallographic experiment must in some manner yield information for the amplitude, frequency and phase of the wave. The amplitude $|\boldsymbol{F}_{hkl}|$ is proportional to the square root of the reflection intensity, and the frequency of each term in the Fourier sum equivalent to the coordinates *h*, *k* and *l* (which are also the spacings of lattices in the crystal in real space). These two components of the wave form are therefore directly measurable in *any* diffraction experiment, satisfying two-thirds of the unknown variables of the electron density equation. The problem arises in the derivation of phases to complete the equation, which is commonly referred to as the *phase problem*. As such, a simple native dataset cannot formulate the complete electron density. This problem is brought about by the two-dimensional nature in which a diffraction pattern is recorded, which essentially 'flattens' the distance measurements from the crystal to the detector, meaning that every reflection ostensibly has the same phase (Figure 1.38). In reality, every reflection has a unique phase (given that it is the summative wave produced by all volume elements along a set of parallel real lattice planes).

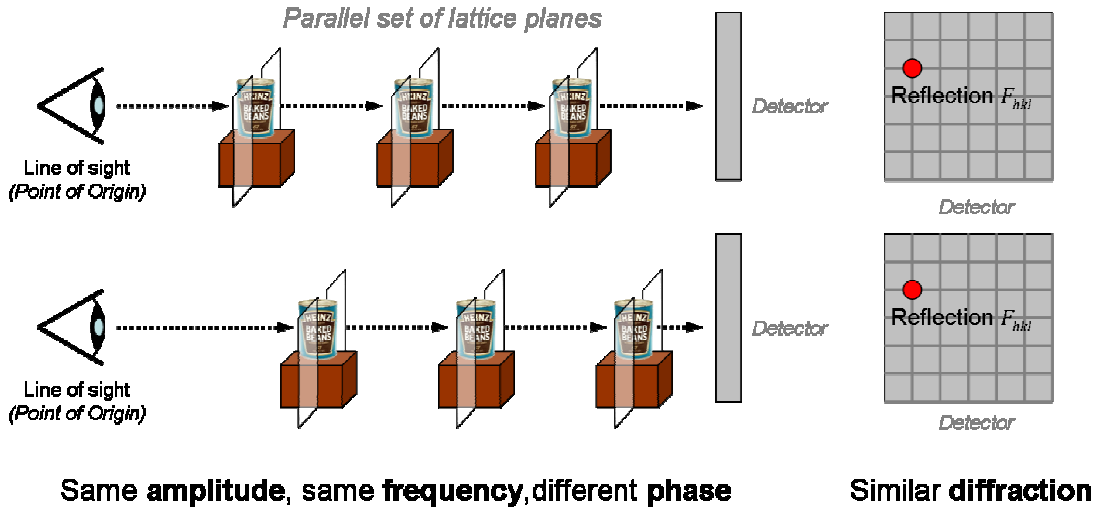


Figure 1.38: **Illustrating the Phase Problem** – In theory, the crystallographic phase problem is equivalent, in practice, to being unable to determine the depth of an object relative to a point of origin, even if we know something about where the object is positioned in terms of height and how many objects are placed in series along a single axis. Regardless of where these objects are positioned along our line of sight, they would, in crystallographic terms, produce the same reflection in a diffraction experiment.

There are two categories of phasing methods, each with their own merits and downsides. As the Protein Databank (PDB) grows, more macromolecular structures are being solved by *molecular replacement* (MR), the process of using an existing atomic model to approximate the phases of an unknown electron density map. The transform of the Fourier sum of the *MR model* can be used to substitute the unknown phases of the new map. These are then iteratively refined until no further improvements can be made to match the calculated wave forms and the observed data. Alternatively, *heavy atoms* can be used for *isomorphous replacement* or *anomalous scatterers* for *single- or multiple-wavelength anomalous dispersion* (SAD/MAD) experiments. Some cases may also be permissive to a combination of both methods (called SIRAS/MIRAS, or *single- or multiple isomorphous replacement with anomalous scattering*). Such atoms can either be incorporated within the expression system (e.g. selenomethionine) or after crystallisation, by soaking existing crystals in a heavy atom solution (such as mercuric chloride or sodium bromide). Heavy atoms may form covalent (e.g. mercury or gold, which target histidines and cysteines) or non-covalent links (e.g. bromide or iodide, which bind glutamate and aspartate residues) with a protein. In both cases, specialised software exists to aid in the derivation of phase information.

1.4.4 Solving the Phase Problem by Molecular Replacement

While the methods to solve the phase problem by MR differ in their application of mathematical principles, the underlying basis for obtaining phase information is the same: use an existing model with the aim of using the phases of this model as a substitute for the unknown phases of the target structure. Phases are borrowed from existing models deposited in the PDB.

There are two main routes to successful phasing by MR: isomorphous and non-isomorphous replacement. If the crystal system of the unknown structure shares the same properties as the known structure (including its point symmetry and space group), and the atoms share a similar location in space (i.e. they are

isomorphous), then the phases of the known structure can be used directly in completing the electron density function of the unknown.

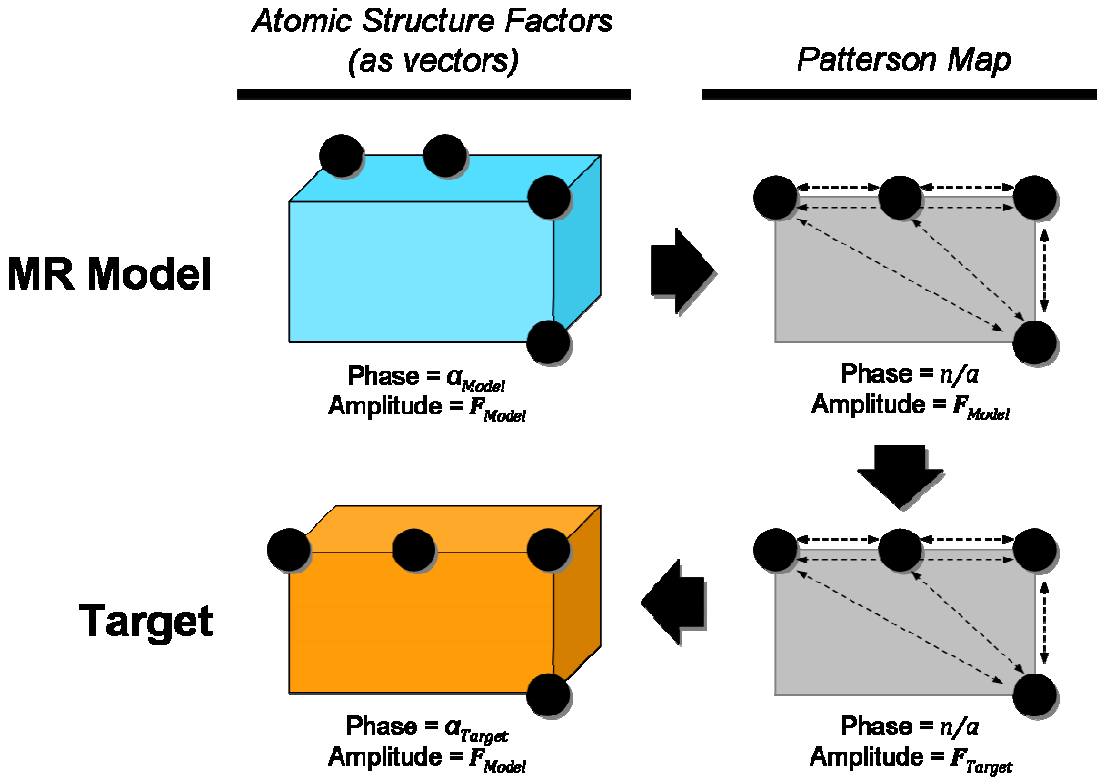


Figure 1.39: **The Use of Patterson Maps in Molecular Replacement** – Any given diffraction pattern can be expressed as a wave function without specification of the phase of the wave. This function is called a Patterson function, and entails the collection of vectors that describes all of the inter-atomic distances in a crystal system. Using Patterson maps (that is, the collection of inter-atomic vectors) from a molecular model and as calculated from a diffraction pattern allows some means toward solving the phase problem computationally, without any experimental intervention. ***NB.*** Each double-headed vector in the Patterson map is a simplified representation of two vectors going in opposite directions.

However, an ostensibly high sequence identity or similar predicted tertiary structure does not assure crystallisation in the same morphology; protein concentration, domain truncations and subtle differences in surface exposed residues mediating crystal contacts will each play a role in the crystallisation process. In the more common method of non-isomorphous MR, the molecular structure factor of the known structure must be expressed in a form that mirrors the real space placement of the unknown structure.

In order to facilitate this, the structure factor functions of the known model are calculated and those of the unknown structure measured, while setting the phase

to 0, meaning that the wave functions of both the MR model and the unknown structure originate from the same arbitrary point in space (i.e. they are *in* phase by virtue of having no phase). The function that describes this is very similar to the electron density function and is effectively the Fourier transform of the measured amplitudes rather than the structure factors:

Eq. 1.9

$$P(u,v,w) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}\left|F_{hkl}\right|^{2} e^{-2\pi i(hu+kv+lw)}$$

This type of function is called a *Patterson function* and is akin to flattening a three-dimensional structure to facilitate a process of 'shape matching'. The flattened structures can be described by a *Patterson map*, which is the collection of vectors that illustrates the inter-atomic distances between volume elements. The aim of molecular replacement is to match the vector expressions for both input functions (Figure 1.39). Since the input structure will have inherent differences to the unknown structure, some of these vectors will be the same, while others will be different. In this respect, molecular replacement is more of a similarity search that functions on a 'best-fit' basis, rather than an identity search.

Two proteins that are related in structure but have been described in different lattice systems, for example, are related by six parameters: three for their relative orientation and three for their relative position in space (one for each of three spatial dimensions). Programs like Molrep (Vagin *et al.*, 1997) and AMoRe (Navaza, 1994) attempt to match the Patterson maps of both the input and target by altering these parameters through a series of *rotations* and *translations* until the two are effectively superimposed (Figure 1.40). Only when these coincide geometrically are the phases of the input model recalculated and utilised in generating a starting electron density map for the unknown structure. In most instances, this geometric coincidence is quite poor because the sheer number of vectors that are needed to describe a large protein molecule will introduce a lot of noise into the Patterson map.

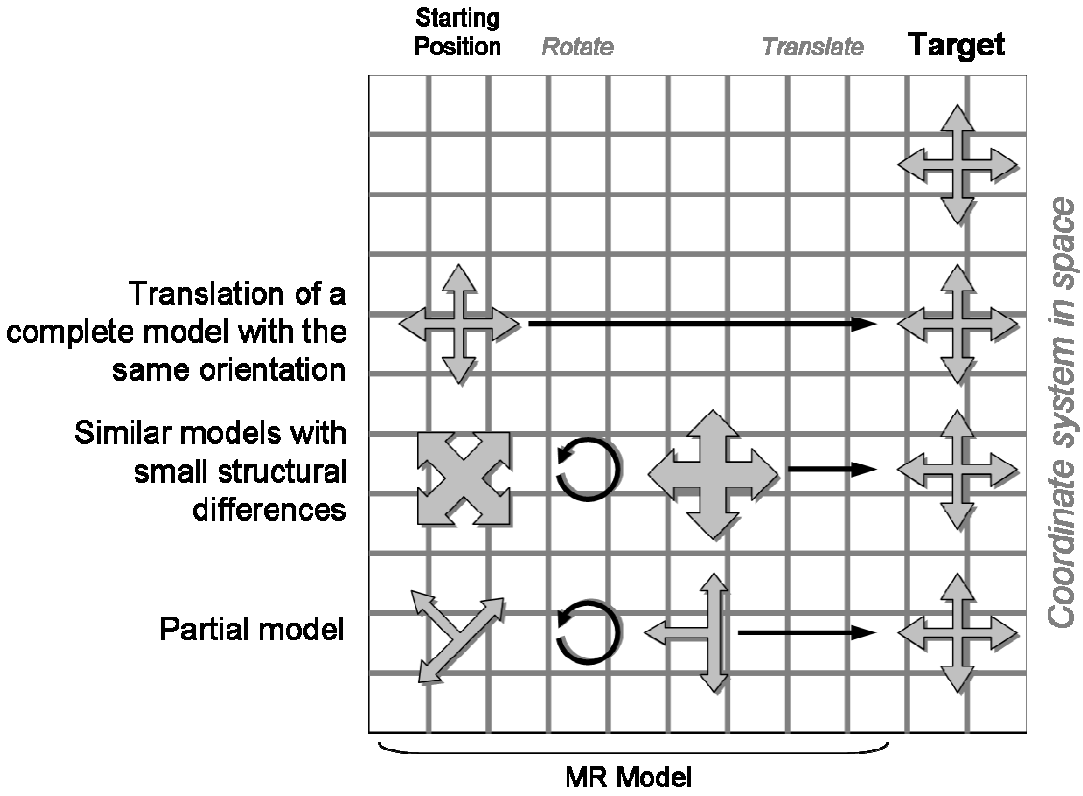


Figure 1.40: **Illustrating the Purpose of Rotation and Translation Functions** – The goal of all molecular replacement programs is to make the Patterson map of the model and the target look similar in space before the phases of the target are substituted for those of the model. Whether the model is structurally identical, has slight structural differences or is a partial model, the goal is to obtain maximise the phasing power of the molecular replacement process in order to obtain a good starting map for the target structure.

Molecular replacement in this manner represents a brute force approach to phasing, since the Patterson function is altered iteratively by a total of six parameters to achieve a match. This becomes computationally intensive in situations where the model is a poor match (with weak similarity at the residue level) and when the contents of the crystal are very complex, such as when multiple copies of different molecules are being searched for.

The program *Phaser* (McCoy *et al.*, 2007) utilises maximum-likelihood statistics to increase the probability of finding an MR solution. Phaser considers the likelihood that the atoms in the input model at any given rotation or translation would produce diffraction data approximating the unknown target (Figure 1.41). Instead of applying a large number of rotational and translational modifiers to every vector of a Patterson map, the maximum-likelihood method takes a more holistic approach to MR by calculating the probability of observing a set of

theoretical reflections from the model that are equivalent to a set of reflections in the diffraction data. The more equivalent reflections are found in each rotation and translation, the greater the likelihood that the model phases will be a reasonable approximation of the target phases. Rotational or translational matches with poor statistical significance over a user-specified threshold are discarded, with the expectancy that every successive term in a complex search (i.e. a molecular replacement search with multiple components) will improve the signal:noise ratio and add to the phasing power of the model.

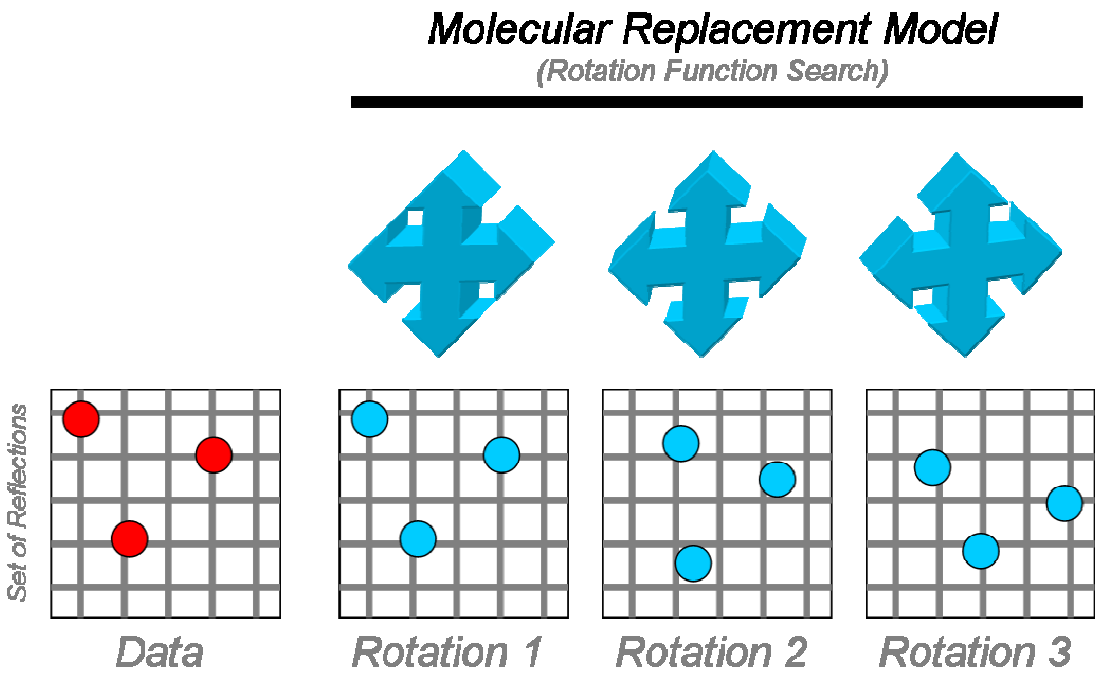


Figure 1.41: **Maximum Likelihood and Molecular Replacement in Phaser** – The maximum-likelihood method attempts to evaluate the theoretical diffraction data of a model against the observed diffraction data from the experiment for a likely match. As the model is rotated and translated within the crystal system specified by the diffraction data, its theoretical diffraction does as well. Only once a proper rotational and translational match has been found, are the phases substituted from model to target in order to derive a starting electron density map.

Phaser assigns every possible solution a Z-score and a log-likelihood gain (LLG). The LLG is a measure of randomness, and essentially measures whether an applied rotation or translation of the model produces a more or less random account of the observed diffraction data. A negative value therefore means that the input search model is worse than a random collection of atoms at describing the target data (a low signal:noise ratio). In a complex search, every step should yield a significant increase in the LLG because the level of randomness should

decrease as more of the data and model match. As many of the search results might have a significant match to some proportion of the data, the Z-score measures the number of standard deviations of these statistical likelihoods above the mean, effectively weighting them in an order of randomness. The greater the Z-score, the more significant the match is.

In this manner, MR allows the crystallographer to solve any structure bearing some degree of homology to a known model, provided the applied strategy is appropriate. One such strategy involves the truncation of all side chains such that each of them resembles either an alanine or a serine, which has the summative effect of removing the noise contribution from a model that has poor sequence identity to the target. Alternatively, a similar approach can be applied while maintaining highly conserved residues between the target and the model, or a family of proteins that share the same structure. In this way, the remaining non-truncated residues of the model can act as 'anchors' and provide some measure of phasing power, which has particular merits when trying to determine a starting point for model building and 'threading' the amino acid sequence of the target protein into an electron density map. This approach is particularly useful for repeat proteins, where multiple iterations of the same secondary structure are repeated in space such that the MR search cannot 'lock on' to the true orientation of the protein in the crystal system because so many aspects of the target itself look the same, and therefore produce similar vectors. If one or more residues can be identified by their distinct electron density, then the power of the MR search – whether by Patterson maps or maximum-likelihood – increases significantly.

If phasing by molecular replacement is not possible, then further experiments can be conducted to facilitate phasing. These methods exploit the unique interaction of atoms with large electron shells to obtain starting phases for model building. As such, these approaches often entail building the model from scratch, though if the phasing process was successful and the resolution is high enough, then the starting map should provide sufficient information for positioning the protein backbone.

1.4.5 Solving the Phase Problem through the Use of Heavy Atoms

Much like MR, experimental phasing techniques exploit isomorphism of datasets to facilitate the process of obtaining starting phases for model building. In doing so, two datasets are needed: one native and one crystal being a *derivative* of the native form but carrying heavy atoms. The term 'heavy atom' relates to the fact that these atoms bear significantly more electrons than other atoms present in the crystal. This added scattering power is critical to assessing the phase differences between the native diffraction pattern and the derivative heavy atom dataset, obtained from derivative crystals.

Commonly used heavy atoms include gold, mercury and platinum, which are all highly toxic substances. Classically, these are introduced into a protein lattice by soaking existing crystals in a solution of heavy atoms in iterative steps over several days until all the binding sites for the heavy atom are saturated. In this process, however, the packing of molecules in the crystal, which could affect the crystal morphology, space group and the resolution, could be changed to such an extent that MIR can no longer be used.

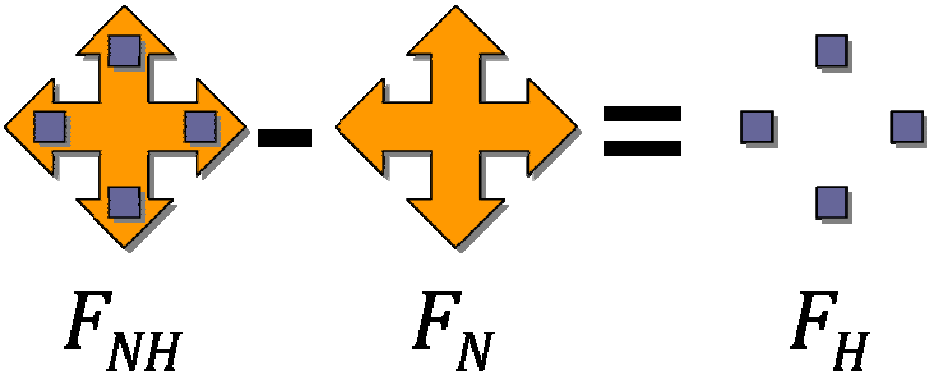


Figure 1.42: **Visualising the Difference Patterson Function** – In an isomorphous replacement experiment, the initial goal is to determine the location of the heavy atoms F_H in space. The difference Patterson map helps in this regard, by removing the noise contributions from the native data (F_N) in the derivative dataset (F_{NH}).

Isomorphous replacement operates on the premise that two crystals remain isomorphous after soaking in a solution of heavy atoms. If this is the case, then two datasets are obtained at the same wavelength and compared in order to

locate the heavy atom in the unit cell. As the structure factor expression for the native (F_N) and derivative heavy atom (F_{HN}) datasets is a summative vector of the scattering from all individual atoms, and the sole difference between the two datasets is the heavy atom signal, subtracting the derivative dataset from the native yields information on the location of the heavy atoms in space (Figure 1.41).

To facilitate phase derivation, both datasets can be expressed as Patterson maps, with the vector sums of these maps equivalent to the electron density function without any assumption of the phase. The difference between these two maps – a *difference Patterson function* (Equation 1.10) – gives the location of the heavy atom. In terms of the electron density, therefore, this method of phasing attempts to obtain a description of the electron density of the heavy atom in space.

Eq. 1.10
$$\Delta P(u,v,w) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}\Delta F^{2}{}_{hkl}\, e^{-2\pi i(hu+kv+lw)}$$

The difference Patterson function is similar to the electron density function of the entire crystal, but is specific to the heavy atom itself and doesn't attribute a position (phase) with respect to an arbitrary origin. One aspect of the Patterson function to bear in mind is that it comprises a set of vectors that describe *all* inter-atomic relationships, meaning that there are actually more vectors in the Patterson map than there are atoms. The benefit of the difference Patterson function is that all of this additional 'noise', attributable to all of the additional distances, is taken away.

At this point, the sets of heavy atoms can be located by probing all of the possible arrangements of atoms with the distances specified by the Patterson map for the correct arrangement of the heavy atoms in a unit cell (Figure 1.43). As the distances *between* atoms are fixed and the Patterson map is centrosymmetric (that is, it can be inverted around its origin), there are two solutions (or two sets of peaks), which fits the description of the atom locations provided by the Patterson map. By tracing the Patterson map from the origin, through an

arbitrary peak, and via the remaining vectors, every possible combination of peaks is computationally investigated for a match to the peaks in the Patterson function. While the determination of the heavy atom coordinates requires three dimensions, the underlying principle is the same.

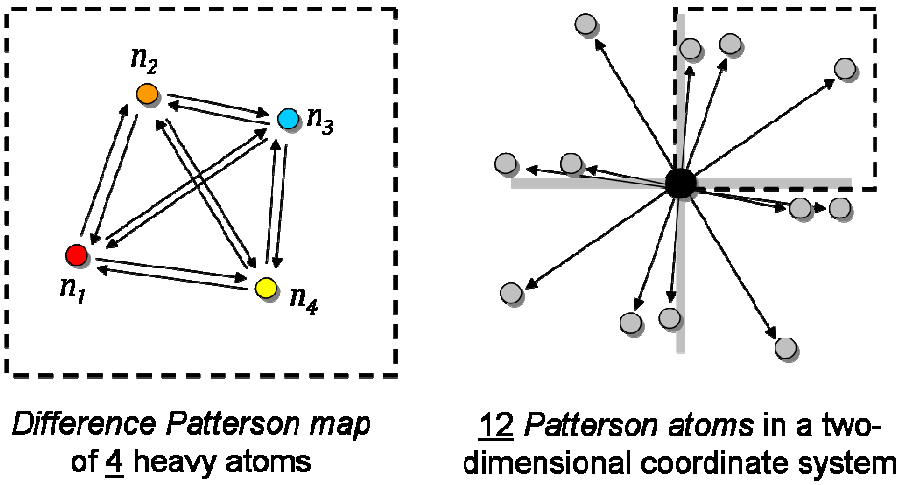


Figure 1.43: **Difference Patterson Map of a Three-Atom Coordinate System** – The vectors that describe the difference Patterson map can be used in context of a simple coordinate system in order to determine the absolute and relative locations of the heavy atoms. By placing each of the vectors with the tail-end on the origin of the coordinate system, all of the possible combinations of atoms can be trialed for the correct arrangement of the heavy atoms in the unit cell. This figure describes the general mathematical principle applied by programs that perform this heavy atom search.

Knowing the location of the heavy atom and its phase, it is possible to determine the phase angle of the vector F_N with the aid of a *Harker diagram.* First, the vector describing the molecular structure factor of the native dataset F_N is plotted in an arbitrary orientation from the origin of a two-dimensional coordinate system, and a circle drawn with a radius equivalent to its length (the amplitude) and with a centre at the end of the vector (Figure 1.44, purple circle). This circle represents all of the potential phase angles of the vector F_N. To this diagram is added the vector specific to the molecular structure factor of the heavy atom (including its phase), with the end of the vector placed at the origin of the coordinate system (Figure 1.44, black arrow). The sum of the heavy atom and the native vector is F_{NH}, which is the structure factor of the derivative dataset bearing the heavy atom (Figure 1.44, red arrow). Therefore, a circle with the radius of this vector will represent all of the possible phases for F_{NH}, which will

satisfy $F_{HN} = F_N + F_H$ (Figure 1.42). The red circle in Figure 1.44 shows how the Harker diagram is completed and highlights the fact that the intersecting points of the purple and red circles are the only two points which satisfy this equation. The phase of the structure factor F_{HN} is therefore indicated by one of the two orientations of the vector F_{HN} with respect to its origin.

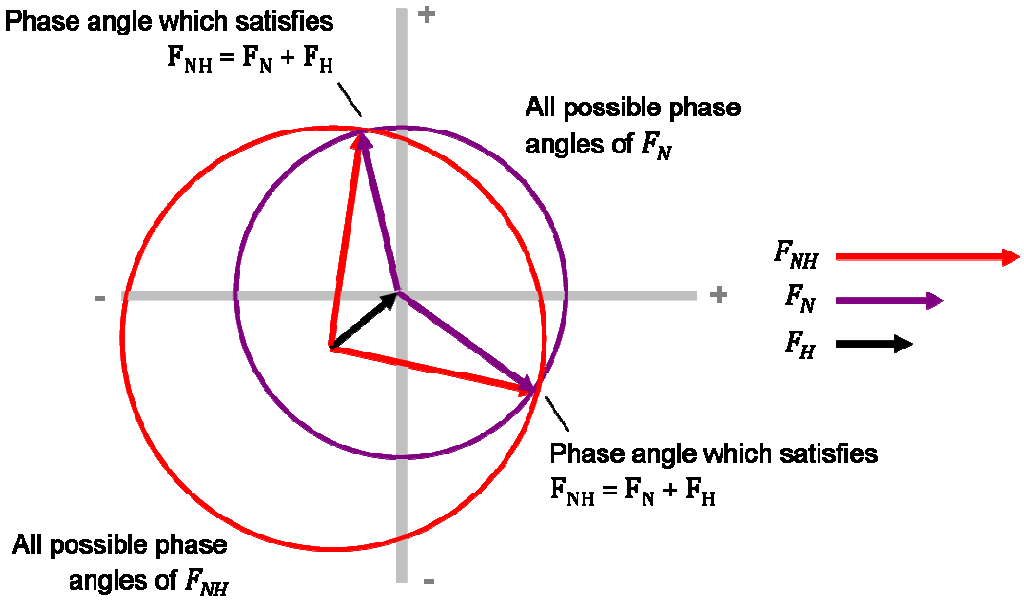


Figure 1.44: **A Harker Diagram for the Determination of Possible Phase Angles for the Vector F_N** – The Harker diagram is a method and simple proof for the derivation of phases from the location of heavy atoms within the unit cell. The circles represent all possible phase angles of the vectors of identical colours (**Red**, for the structure factor of the derivative dataset, and **Purple** for the native). Knowing the phase and amplitude of the vector F_H, and the vector length of the derivative data and native dataset (F_{NH} and F_N, respectively) allows the identification of the phase angle of the native structure factor as one of two possible solutions. In order to definitively identify this phase, a second derivative dataset can be obtained; alternatively, two maps using both phase solutions can be compared qualitatively, since one is likely to provide a better starting map than the other.

Typically, in order to determine which is the correct phase, a second derivative dataset must first be obtained and the same process of locating the heavy atom to determine its phase repeated. When this second heavy atom vector is added to the Harker diagram and similar circles are plotted to provide an estimate of the phase of vector F_{HN}, one of the two intersecting points will match (or closely intersect with) one of the two points from the original phase derivation. At this point, the phase of the native data set can be derived because $F_N = F_{HN} - F_H$.

As an alternative to isomorphous replacement the crystallographer can make use of the unique absorption characteristics of atoms such as zinc, manganese or selenium, which produce an anomalous scattering signal at wavelengths accessible via a synchrotron X-ray source. These wavelengths coincide with the energy requirement to cause an electronic transition in the anomalous scatterer, causing small changes in the scattering intensity.

The most common method involves the substitution of native methionines with s*elenomethionine* during recombinant protein expression. Selenomethionine bears a similar structure to methionine but instead of a terminal sulphur group, contains a selenium group with a heavier scattering signal. Since these types of phasing methods are dependent on the magnitude of the anomalous scattering signal, a threshold number of methionines must be substituted to provide sufficient phasing power.

An *anomalous dispersion* experiment is typically carried out at a wavelength corresponding to the specific *absorption edge* of the anomalous scatterer, where absorption happens more readily than at other wavelengths but part of the X-ray is still emitted with a different phase, at the absorption *peak*, where most of the X-ray is absorbed, and at a *remote* wavelength, where little is absorbed. This type of experiment, utilising multiple wavelengths to alter the diffraction of the crystal, is called *multiple-wavelength anomalous dispersion* (MAD). An evolution of this technique makes use of just a single wavelength (MAD, for *single-wavelength anomalous dispersion*), which is often sufficient to phase a map provided the anomalous signal is well recorded.

During normal scattering, two symmetry-related reflections should have the same intensity such that $\left|\boldsymbol{F}_{hkl}\right| = \left|\boldsymbol{F}_{\overline{hkl}}\right|$, where $\left|\boldsymbol{F}_{hkl}\right|$ is the absolute amplitude $\boldsymbol{F}$ of the set of reflections produced by the planes (*hkl*) with indices *hkl* and $\overline{hkl}$ (Figure 1.44). These reflections are related by the opposite phase (i.e. 360° - α_{hkl}) such that $\alpha_{hkl} = -\alpha_{hkl}$. This is caused by the inherent centrosymmetry in the electron distribution of an atom, which also produces a centrosymmetric

diffraction pattern. SAD and MAD experiments take advantage of the loss of centrosymmetry that results from anomalous scattering, because it causes two symmetry-related reflections to no longer be in phase and induces subtle but measureable differences in their intensity. Instead of a normal scattering factor, f_0, the atomic scattering, f, by an anomalous scatterer is the combination of normal scattering, the absorption that reduces this normal scattering, f', and the factor of anomalous scattering, f'', which is considered *out of phase* with the normal scattering component (and is therefore represented as an imaginary component, *i)* (Equation 1.11).

Eq. 1.11 $$f = f_0 + f' + if''$$

The native amplitude and phase relationship is called *Friedel's law* and two reflections that can be described by this law are part of a *Friedel pair* (Figure 1.45, left). The effect of the anomalous scattering signal is that this law is broken, and the relationship between the two reflections no longer holds true, resulting in a change in the phase and amplitudes of F_{hkl} and $F_{\overline{hkl}}$ (Figure 1.45, right). As the wavelength used in such experiments is fixed by the type of atom that is being use as an anomalous scatterer, the magnitude of the anomalous signal is constant and can therefore be read from reference tables detailing this information for all atoms.

SAD and MAD phasing techniques take advantage of this discrepancy to obtain starting phases. This process is essentially begun in the same manner as in isomorphous replacement, through the generation of a Patterson difference map to locate the origin of the anomalous scattering signal. Knowing the change in the phase of the former Friedel pair based on existing knowledge of the absorption magnitude of the anomalous scatterer, permits the calculation of approximate phases of the original set of reflections, and therefore the positions of all atoms that contribute to those reflections. In this way, starting phase information can yield a significant electron density map into which the starting model can be built.

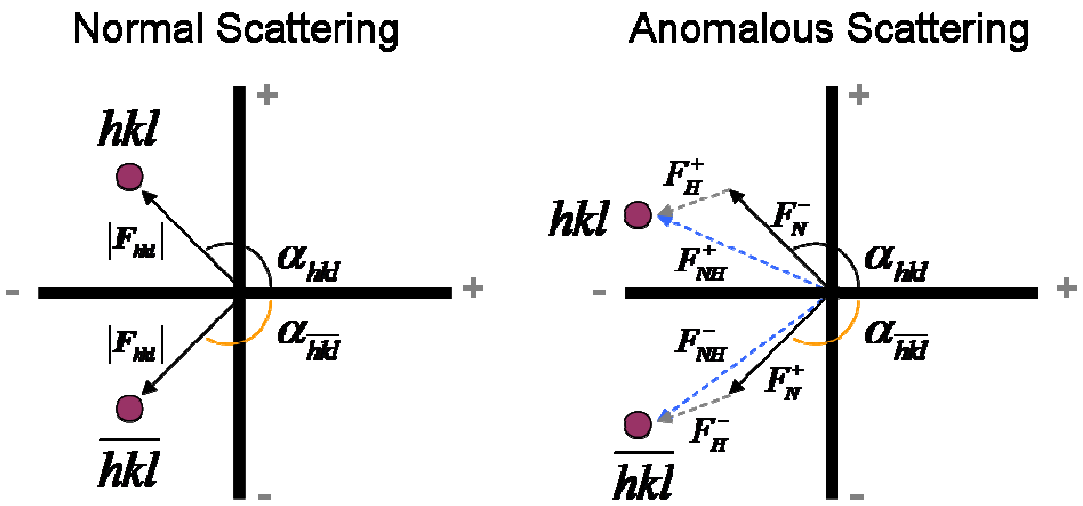


Figure 1.45: **Friedel Pairs** – Friedel's law states that pairs of symmetry-related reflections will share the amplitude $|\mathbf{F}_{hkl}|$ and have opposite phase, (α and 360° - α) (**Left**). However, in an anomalous scattering experiment (**Right**), this effect is broken, resulting in a difference in phase and amplitude of the scattered wave from both parts of the Friedel pair. The anomalous signal is represented as the addition of vector $\mathbf{F}_H$ to the normal scattering vector $\mathbf{F}_N$, resulting the vector $\mathbf{F}_{NH}$. The two reflections, by virtue of being part of a Friedel pair, are centrosymmetric under normal scattering conditions.

1.4.6 Refinement of Macromolecular Models

Once the electron density map has been phased, the goal is to improve the function in iterative steps such that it more accurately describes the protein molecules within the crystal. This is because, whether through experimental methods or molecular replacement, the starting phases for the electron density are either biased or inaccurate. This iterative process is called *structure refinement* and numerous pieces of software exist that employ different methods toward the goal of extracting as much information as possible out of the diffraction data.

The mathematical aim for refinement is to minimise the differences between the observed and the calculated structure factors. All atoms have a position in space (x_j, yj, z_j) and a natural oscillation that is determined by the local environment of that atom. This oscillation is denoted by the *temperature factor,* **B** (or *B-factor*), and it is assumed that main chain atoms will exhibit less freedom of movement than side chain atoms. In macromolecular diffraction experiments, the differences in atomic oscillations causes measureable differences in the

diffraction intensity, due to the inherent variability of atomic positions. Refinement programs will therefore vary four parameters of the calculated structure factors in order to achieve a minimised electron density function.

As the phases of the first map are, in effect, borrowed from the experiments described in Section 1.4.5, the Fourier sum that describes the electron density takes on a different form that accounts for this initial phase information:

Eq. 1.12 $$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} w_{hkl} \left|\boldsymbol{F_{obs}}\right| e^{-2\pi i(hx+ky+lz-\alpha'_{calc})}$$

Where $\left|\boldsymbol{F_{obs}}\right|$ are the observed absolute amplitudes calculated from the intensities of the reflections and α'_{calc} the phases calculated from one of the experiments described in Section 1.4.5. As the reliability of the phases will differ depending on the experiment used to obtain them (for example, how reliable the molecular replacement model is for describing the target structure), the phase of this electron density equation is weighted by a factor, w_{hkl}, reflective of their accuracy. This weighting means that potential bias from a bad molecular replacement model or poor calculation of heavy atom phases can be negated with a low weighting between 0.1-0.5, and will consequently have little affect on the initial calculation of the electron density map (also called the $\mathbf{F_0}$ map). Therefore, the calculation of a starting map is reliant on the accuracy of the phasing experiment that precedes this step in structure determination.

Using molecular modelling programs such as Coot (Emsley *et al.*, 2010), a starting electron density map must first be assessed for quality indicators of an interpretable map. The hallmarks of a good starting map are areas of continuous tubular density that correspond to discernable secondary structure, and in some cases, evidence of side chain density. Evidence for the position of large hydrophobic side chains often serve as an anchor point from which to construct the model (Figure 1.46).

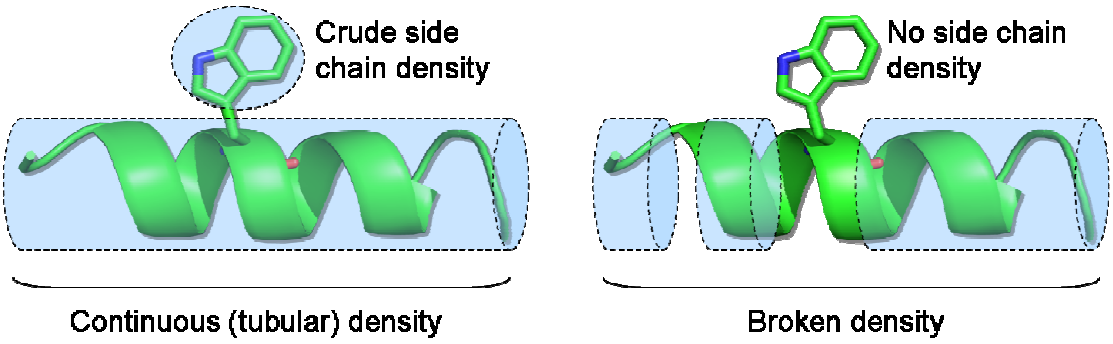


Figure 1.46: **Evaluating the Quality of an Electron Density Map** – The starting electron density map will often contain partial information on the peptide backbone, such as continuous tubular density with no defined shape indicative of a alpha helical secondary structure. Nonetheless, such maps are often a good starting point for building the molecular model. Similarly, side chain densities will often be incomplete until phases are refined. This image was generated from the structure described in Chapter 4.

If the map is deemed to be of reasonable quality, steps can be taken to increase the signal:noise ratio; that is, to improve the contribution of the scattering mass (the protein, nucleic acid or small molecule) and remove the noise contribution from bulk solvent. This process of *density modification* or *solvent flattening* is carried out by programs such as Pirate (Cowtan, 2000), Parrot (Zhang *et al.*, 1997) or DM, and is often useful if the first map looks like a large web of density, rather than distinct tubes of density in space. All three programs utilise existing mathematical knowledge on the behaviour of proteins (or nucleic acids) in relation to solvent to accurately model the areas that are presumed to make up part of the scattering mass. The summative effect of these programs is the smoothing of areas that have significant density and the removal of areas with low or no density, so that continuous regions appear more pronounced, rather than just seeming as jagged shapes in space. In turn, regions which are closely related and could be continuous are joined together into a smooth contoured map. This makes the map more interpretable and means that regions of defined secondary structure are more readily identifiable.

Another method of improving the map during refinement is to maintain the inherent identity of molecules that are related by their structure, but are not in themselves related by some crystallographic symmetry operator. Molecules that are related chemically in this way are related in refinement terms by the aforementioned non-crystallographic symmetry (NCS). NCS restraints impose a

type of averaging that will constrain areas of density that are similar, but will allow other regions to differ. As such, it can be modified to accommodate conformational differences between molecules of the asymmetric unit, and hold identities in structure where appropriate. Alternatively, the unsolved model might be constrained to look like a target structure, if it is assumed that the two share a significant degree of structural and/or sequence homology. Both NCS and these *local structural similarity restraints* (LSSR) are employed by the program BUSTER (Bricogne, 1993) in refining the electron density equation.

The programs that compute the refinement process use one of three main methods to minimise the electron density function: maximum-likelihood, least squares fitting or Bayesian probability methods. Both Refmac (Murshudov *et al.*, 2011) and the aforementioned program BUSTER use maximum-likelihood in order to determine how well the calculated structure factors of the atomic model describe the data. One added benefit of BUSTER is that it can model potential areas of protein density and mark these areas with placeholders (i.e. default objects which represent some scattering mass), allowing more regions of the model to be constructed with each round of refinement. These placeholders, which must be described as an atom in real space, are modelled as low resolution entities in order to mimise the bias of phase information provided by these extra regions. Each refinement program will also consider the chemistry of a macromolecule by imposing geometric constraints on the residues that compose the model and the likely distances that exist between them. In this way, areas that are energetically unfavourable (such as side chains that are clashing or poor helical geometries) are corrected throughout the refinement process as well, provided the new geometry gives a better estimate of the observed structure factors.

Regardless of the program that is used to compute the steps in refinement, each one relies on a quality indicator to assess the accuracy of the molecular model in relation to the data extractable from the diffraction data. This parameter is known as the *R-factor*, which is a measurement of the agreement between the

observed structure factors from the set of reflections and the calculated structure factors from the model:

Eq. 1.13
$$R = \frac{\sum \left| |F_{obs}| - |F_{calc}| \right|}{\sum |F_{obs}|}$$

where F_{obs} are the observed and F_{calc} the calculated structure factors. Using this equation, an R-factor of 63% suggests that the signal:noise ratio of the input model (whether partial or whole) is not any more significant than a random model. There are two types of R-factors that measure the progress of refinement: a factor termed R_{Work} and another called R_{Free} (Brünger, 1992). The R_{Work} is essentially a quantitative *and* qualitative indicator for the entire set of reflections, while the R_{Free} is merely a *qualitative* measure of agreement between the structure factors from the model and a reserved set of reflections that are not subject to the refinement process. The reflections used in the calculation of the R_{Free} must comprise only a small subset of the total (typically 5%) so that enough information remains to be modelled in the refinement process. The purpose of the R_{Work} is to give some sense of the agreement between the diffraction data and the model, with a low R_{Work} reflecting that a large percentage of the diffraction data can be calculated back by the atomic model (or, conversely, that a low percentage of the target function remains undescribed). The role of the R_{Free} measurement is to provide an indication of the level of model bias as more and more phase parameters are added to the electron density equation, since the set of selected reflections remains unaffected by the refinement process. Ideally, the two R-factors will converge and stabilise within 5% of each other at the end of refinement.

The basic premise of refinement, then, is to iteratively return to the electron density equation and renew the phase information. In practice, this means that ideally every successive electron density map contains new information for the location of atoms in the atomic model. As the phases become more accurate, the location of side chains and loop regions becomes clearer, provided they are stabilised within the protein (flexible regions will tend not to produce any

significant form of diffraction). Only when the R-factors have stabilised and no new electron density emerges with additional rounds of refinement is the model said to be finished and ready for deposition in the PDB.

CHAPTER 2: MATERIALS AND METHODS

2.1 GENERAL CLONING PROTOCOLS

LB and TB media were prepared by adding the appropriate weight of dry powder (Merck) to the corresponding volume of distilled water (25.0 g of dry LB powder per litre or 47.0 g of TB powder per litre of water).

Agar plates were prepared using 37.0 g of dry LB agar powder (Sigma-Aldrich) per litre of distilled water and the appropriate antibiotic (final concentration of 100 μg/mL of ampicillin, 50 μg/mL of kanamycin or 34 μg/mL of chloramphenicol) before pouring.

2.1.1 PLASMID PURIFICATION AND QUANTITATION OF DNA CONCENTRATION

Plasmid stocks were prepared from 5-10 mL of *E. coli* TOP10 cultures grown overnight in LB. DNA purification was conducted using the QIAprep Spin MiniPrep kit (Qiagen), and the protocol followed as specified in the user manual for 5 mL cultures. For a 10 mL culture volume, the volumes of buffers required for lysis and precipitation of chromosomal DNA were doubled.

Verification of DNA yield from plasmid purification, PCR or other methods was performed using the 'Nucleic Acid' sub-program in the NanoDrop 8000 (Thermo Scientific) software supplied with a NanoDrop 1000 spectrophotometer (Thermo Scientific). Concentrations were recorded in μg/μl using the elution buffer supplied with the Qiagen MiniPrep kit as a blank for measurements.

2.1.2 PREPARATION OF COMPETENT CELLS FOR CLONING AND EXPRESSION/PURIFICATION

The complete list of competent cells maintained for expression and/or plasmid amplification were as follows:

- ***BL21 Star™ (DE3)*** (Invitrogen): For high level expression of recombinant proteins from low copy number plasmids, with internal mutations that minimise mRNA degradation.
- ***BL21 Star™ (DE3) pRARE:*** Same as BL21 STAR™ but transformed with the rare codon plasmid, pRARE.
- ***BL21 (DE3) pLysS:*** Genotypically identical to BL21 STAR™ but transformed with a plasmid carrying a lysozyme mutant for enhanced suppression of leaky expression.
- ***T7 Express LysY*** (NEB): A BL21 derivative carrying a chromosomally-integrated lysozyme variant with no amidase activity, therefore preventing lysis during induction.
- ***T7 Express LysY/I^{q}*** (NEB): Same as T7 Express LysY but with greater control of expression via the *lacI^{q}* gene.
- ***One Shot® TOP10*** (Invitrogen): Specialised strain of *E. coli* for stable replication of high copy number plasmids, modified for high efficiency transformation.

Competent cells were prepared through a protocol adapted from Hanahan, 1985 and Hanahan, 1983. The following buffers were used to promote or restore competency:

- **TFB1**: 30 mM potassium acetate, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 100 mM RbCl, 15% glycerol (filter sterilised and stored at room temperature).
- **TFB2**: 100 mM MOPS (pH 6.5), 75 mM $CaCl_2$, 10 mM RbCl, 15% glycerol (filter sterilised and stored at 4°C).

Briefly, 2.5 mL of an overnight culture of cells, prepared from a single colony (transformed or not), was added to 250 mL of autoclaved LB media supplemented with 20 mM $MgSO_4$ in a sterile manner to an OD_{600} of 0.4-0.6. Cells were subsequently pelleted by centrifugation at 3500 x g for 5 minutes at 4°C in 50 mL Falcon tubes (Fisher Scientific) pre-cooled on ice. The supernatant from this step was decanted and the cell pellet gently resuspended by pipetting in 0.4x the original volume of culture in TFB1, while keeping everything on ice. After 5

minute incubation, the cell suspension was again pelleted by centrifugation at 3500 x g for 5 minutes and the supernatant gently aspirated by pipetting, before adding 1/25th of the original cell culture volume of TFB2 to this pellet and resuspending gently by pipetting. The cell suspension was then incubated on ice for at least one hour before aliquoting into 50 µL aliquots and flash freezing in liquid nitrogen. Cells were stored at -80°C until required for transformation (Section 2.1.7).

2.1.3 GENE AMPLIFICATION VIA THE POLYMERASE CHAIN REACTION (PCR)

For cloning, the polymerase chain reaction (PCR) was used to amplify genes or gene fragments from existing plasmids, or for diagnostic purposes to verify cloning procedures described in Section 2.1.4 and 2.1.5 (in particular for smaller genes). For high fidelity *de novo* gene amplification one of either Herculase II Fusion, Pfu Turbo or Pfu Ultra II polymerase (Agilent Technologies) were added in volumes specified by the manufacturer. The dNTP mix (Promega) was prepared as a 1:10 dilution from a 100 mM stock, while the plasmid template was typically prepared as a 1:10 dilution from a purified plasmid stock of any given concentration. Appropriate forward and reverse primers were designed and synthesised by Sigma-Aldrich or Eurofin MWG Operon; the complete list of primers used to generate constructs for this study are listed in Table 2.1, with the exception of those that were given as gifts.

Primer	T_m (°)	Sequence
Standard Amplification Primers		
Skp1_BTBPOZ_for	69.5	GTG ACT TCT AAT GTT GTC CTA GTG AGT GG
Skp1_BTBPOZ_rev	70.9	CTA GAT CAT TTC AGC AAC CAC CTT GC
Mutagenesis Primers		
Sgt1_L48R_sense	73.6	CTC ATT TTT AAG GCA GCA TGT CGT GAA AAG CTA TAC TTT GGG TTT
Sgt1_L48R_anti		AAA CCC AAA GTA TAG CTT TTC ACG ACA TGC TGC CTT AAA AAT GAG
Sgt1_W58A_sense	72.4	AGC TAT ACT TTG GGT TTT CTG ATG CGC ACA GCG ATG CAA CT
Sgt1_W58A_anti		AGT TGC ATC GCT GTG CGC ATC AGA AAA CCC AAA GTA TAG CT
Sgt1_H59A_sense	72.5	CTT TGG GTT TTC TGA TTG GGC CAG CGA TGC AAC TAT GGA A
Sgt1_H59A_anti		TTC CAT AGT TGC ATC GCT GGC CCA ATC AGA AAA CCC AAA G
Sgt1_D61R_sense	72.5	GTT TTC TGA TTG GCA CAG CCG TGC AAC TAT GGA AAA CGC A
Sgt1_D61R_anti		TGC GTT TTC CAT AGT TGC ACG GCT GTG CCA ATC AGA AAA C
Sgt1_L216A_sense	74.6	CGT TGA TGA CAC TTT GCC TGC ATG GGA AGA TCG TCT GGA G
Sgt1_L126A_anti		CTC CAG ACG ATC TTC CCA TGC AGG CAA AGT GTC ATC AAC G
Sgt1_W127A_sense	73.4	GTT GAT GAC ACT TTG CCT TTA GCG GAA GAT CGT CTG GAG AC
Sgt1_ W127A _anti		GTC TCC AGA CGA TCT TCC GCT AAA GGC AAA GTG TCA TCA AC
Sgt1_R93A_sense	70.1	ATA GAT CGA AGA TTG GTC TTG TTA ACT TTG CTT ATT TTG TTC ATT TCT TCA ATA TCA AAG
Sgt1_R93A_anti		CTT TGA TAT TGA AGA AAT GAA CAA AAT AAG CAA AGT TAA CAA GAC CAA TCT TCG ATC TAT

Table 2.1: **List of Primers for Standard PCR and Mutagenesis**

Each PCR reaction contained the following standard reagents and was made up to 50 µL using Milli-Q water (Millipore):

Reagent	Final Concentration
Template DNA	0.1-0.4 µg/µL
dNTP Mix (Promega)	20-40 µM
DNA polymerase (Agilent Technologies)	0.5-2.0 units
Forward primer	200 nM
Reverse primer	200 nM
Polymerase buffer	1 X
DMSO	2% (v/v)

Table 2.2: **Composition of PCR Reactions**

PCR reactions were carried out using a SureCycler 8800 Thermal Cycler (Agilent Technologies) with the following standard program:

1) Heated lid and hot start at 96°C
2) Denaturation at 96°C (1 minute)
3) Annealing at a temperature within 5°C of the primer melting temperature (T_m) (1 minute)
4) Elongation and amplification at 72°C (30-60 seconds per kilobase of DNA being amplified, depending also on the polymerase being used)
5) 10 minutes at 72°C

Steps 2-4 were cycled 15-20 times for sufficient amplification. Successful amplification was confirmed visually using a Dark Reader blue light transilluminator (Clare Chemical Research).

2.1.4 Site-Directed Mutagenesis by PCR

General procedure for PCR mutagenesis followed the protocol outlined by the QuikChange Site Directed mutagenesis kit (Agilent Technologies), using 12-15 cycles of amplification, instead of 15-20 for standard gene amplification. Mutagenesis primers were designed using the QuikChange Primer Design tool on the Agilent Technologies website and are outlined in Table 2.1.

2.1.5 RESTRICTION-DIGESTION CLONING

Restriction enzyme-digested genes were prepared for directional sub-cloning into recipient expression vectors. The following general reaction mixture was set up for each preparative digest:

Reagent	Volume (µL)
5' restriction enzyme (NEB or Roche)	1
3' restriction enzyme (NEB or Roche)	1
Optimal restriction buffer (specified by the NEB 'Double Digest Finder' tool [http://www.neb.com/nebecomm/doubledigestcalculator.asp])	3
10X BSA stock (prepared from 100X NEB-supplied BSA stock)	3
Vector (with or without gene of interest) or PCR product (prepared as described in Section 2.1.1)	10
Filter sterilised Milli-Q water (Millipore)	12

Table 2.3: **Composition of Restriction Digest Reactions**

BSA was always added as standard to each reaction mixture. All reactions were incubated at 37°C with shaking for 1-2 hours for a PCR product or existing gene (insert), and 3-4 hours for a recipient vector. Digested vectors and inserts were then purified using an Illustra GFX PCR DNA and Gel Band Purification kit (GE Healthcare), according to the manufacturer's instructions.

2.1.6 LIGATION

For ligation, restriction enzyme-digested recipient vector, and PCR products or excised genes (insert), were run on a 0.75-1.5% agarose gel and gel purified as outlined in the Illustra GFX PCR DNA and Gel Band Purification kit (GE Healthcare). Recipient vector DNA was diluted 10-fold prior to use. The following reaction mixes were prepared for ligation, for a total reaction volume of 10 µL:

Reagent	Volume (µL)
Recipient Vector	1
Insert	1, 2 or 4 (for 1:1, 2:1 or 4:1 ratio of insert:vector)
T4 DNA Ligase (NEB)	1
T4 DNA Ligase buffer (NEB)	1
Milli-Q Water	6, 5 or 3 (depending on insert:vector ratio)

Table 2.4: **Composition of Ligation Reactions**

In addition, a control ligation was set up for each ligation where no insert was present, to determine the level of background ligation occurring due to improperly digested vector. All reactions were prepared in autoclaved 1.5 mL tubes (Eppendorff). Reaction mixes were left at room temperature for 20-180 minutes and then transferred onto ice to cool the reaction, or to -20°C for storage until needed.

2.1.7 Transformation of Chemically Competent Cells by Heat Shock

Transformation of chemically competent *E. coli* strains for expression or DNA plasmid amplification was conducted by heat shock. 50-100 µL of competent cells were thawed on ice. After thawing, 1 µL of recombinant plasmid DNA from the selection listed in Table 2.5 was added to the competent cells and the mixture incubated on ice for at least 15 minutes and no more than 60 minutes. Cells were then rapidly transferred to a heating block maintained at 42°C for 45 seconds, then transferred back to ice for 2 minutes. 500 µL of sterile autoclaved LB media was added to this heat shocked mix and placed at 37°C for at least one hour (no more than 2 hours) with shaking. Recovered cells were then spread onto pre-prepared agar plates containing the appropriate antibiotic (Section 2.1). Plates were incubated either at 37°C to be used the following day, or at room temperature for a maximum of 2 days.

2.2 Recombinant Protein Expression Protocols

Expression of recombinant gene products was generally conducted in specialised *E. coli* cells grown in LB or TB media (Merck), as detailed in Table 2.5. Media was prepared as described in Section 2.1.

Agar plates used to obtain single colonies for preparation of overnight cultures for protein expression were prepared as described in Section 2.1.

2.2.1 SMALL-SCALE EXPRESSION (EXPRESSION STUDIES)

Bacterial cultures of less than 200 mL volume were used to optimise expression conditions (temperature, cell line, etc.) of genes that had not been previously characterised.

10-200 mL of autoclaved LB media with the appropriate antibiotic was inoculated with 1/100th the volume of the expression culture using a small pre-culture of cells transformed with the plasmid under investigation and grown overnight. All studies involved an initial growth phase at 37°C before being transferred to different temperatures for expression. Gene expression at an OD_{600} between 0.6-0.8 was induced with a final IPTG concentration of 1 mM. Cultures were supplemented with 50% of the volume of antibiotic initially added to the culture. Protein expression was carried out at 37°C for 3 hours or 18°C overnight.

Pre- and post-induction samples were prepared in 3X gel loading buffer (GLB) (100 mM Tris pH 6.8, 30% glycerol, 3% SDS, 0.01% bromophenol blue, 350 mM DTT) for analysis by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) (Section 2.4).

In instances where protein solubility was also being investigated, the complete culture was centrifuged at 4500 x g for 10 minutes at 4°C. After resuspension of the cell pellet in 1/10th the culture volume and lysis of cells, the soluble fraction was obtained by centrifugation of the lysate at 13000 RPM using a FA-45-18-11 fixed angle rotor (Eppendorf). 'Soluble' and 'whole cell' fractions were prepared for analysis by SDS-PAGE (Section 2.4).

Plasmid Name	Gene	Tag	Vector	Resistance	Origin of Replication	Cloning Sites	Expres. Strain
Sgt1 Constructs							
Sgt1 WT-Untagged-pCOLADuet-1	Full length Wild Type Sgt1	-	pCOLADuet-1 (MCS 2)	Kanamycin	ColA	-	BL21*
Sgt1 WT-SII-pCOLADuet-1	Full Length Wild Type Sgt1	C-terminal Strep-tag	pCOLADuet-1 (MCS 2)	Kanamycin	ColA	-	BL21*
Sgt1 WT-His-pET28a	Full Length Wild Type Sgt1	N-terminal His_6-tag	pET28a	Kanamycin	pBR322	NdeI:XhoI	T7 Express LysY/I^Q
Sgt1ΔSGS-SII-pCOLADuet-1	Sgt1 Wild Type, Residues 1-279	C-terminal Strep-tag	pCOLADuet-1 (MCS 2)	Kanamycin	ColA	-	BL21*
Sgt1 1-178-Untagged-pRSETa	Sgt1 Wild Type, Residues 1-178	-	pRSETa	Ampicillin	pUC	NdeI:BamHI	T7 Express LysY/I^Q
Sgt1 1-150-Untagged-pRSETa	Sgt1 Wild Type, Residues 1-150	-	pRSETa	Ampicillin	pUC	NdeI:BamHI	T7 Express LysY/I^Q
Sgt1 1-136-Untagged-pRSETa	Sgt1 Wild Type, Residues 1-136	-	pRSETa	Ampicillin	pUC	NdeI:BamHI	T7 Express LysY/I^Q
Sgt1 1-121-Untagged-pRSETa	Sgt1 Wild Type, Residues 1-121	-	pRSETa	Ampicillin	pUC	NdeI:BamHI	-
Sgt1 L48R-His	Full Length Sgt1 Mutant L48R	N-terminal His_6-tag	pET28a	Kanamycin	pBR322	NdeI:XhoI	T7 Express LysY/I^Q
Sgt1 H59A-His	Full Length Sgt1 Mutant, H59A	N-terminal His_6-tag	pET28a	Kanamycin	pBR322	NdeI:XhoI	T7 Express LysY/I^Q
Sgt1 W58A-His	Full Length Sgt1 Mutant, W58A	N-terminal His_6-tag	pET28a	Kanamycin	pBR322	NdeI:XhoI	-
Sgt1 D61R-His	Full Length Sgt1 Mutant, D61R	N-terminal His_6-tag	pET28a	Kanamycin	pBR322	NdeI:XhoI	T7 Express LysY/I^Q
Sgt1 R93A-His	Full Length Sgt1 Mutant, R93A	N-terminal His_6-tag	pET28a	Kanamycin	pBR322	NdeI:XhoI	T7 Express LysY/I^Q
Sgt1 L126A-His	Full Length Sgt1 Mutant, L126A	N-terminal His_6-tag	pET28a	Kanamycin	pBR322	NdeI:XhoI	T7 Express LysY/I^Q
Sgt1 W127A-His	Full Length Sgt1 Mutant, W127A	N-terminal His_6-tag	pET28a	Kanamycin	pBR322	NdeI:XhoI	T7 Express LysY/I^Q
Skp1 Constructs							
Skp1 WT-Untagged-pET22b	Full Length Wild Type Skp1	-	pET22b	Ampicillin	-	-	BL21* pRARE
Skp1Δ BTB/POZ-Untagged	Skp1 Wild Type, Residues 1-158 (Deletion between 32-74)	-	pRSETa	Ampicillin	pUC	NdeI:BamHI	T7 Express LysY/I^Q
Skp1 BTB/POZ-Untagged-pRSETa	Skp1 Wild Type, Residues 1-158	-	pRSETa	Ampicillin	pUC	NdeI:BamHI	T7 Express LysY/I^Q
Skp1 BTB/POZ-His-pET28a	Skp1 Wild Type, Residues 1-158	N-terminal His_6-tag	pET28a	Kanamycin	pBR322	NdeI:BamHI	T7 Express LysY/I^Q

Table 2.5: **Summary Table of Standard Plasmids Used for Recombinant Protein Expression** – 'Sgt1 WT-Untagged-pCOLADuet-1', 'Sgt1 WT-SII-pCOLADuet-1' and 'Sgt1ΔSGS-SII-pCOLADuet-1' were originally cloned by Cara K Vaughan. 'Skp1 WT-Untagged-pET22b' was originally cloned by Alan Purvis. All of these constructs were given as gifts in aid of this study.

2.2.2 Large-scale Cell Culture Growth for Recombinant Protein Expression

Single colonies from fresh transformations were picked into 100 mL of LB media and incubated overnight at 37°C. 10 mL of this pre-culture was added to each 1 L of LB media and cells grown at 37°C to an OD_{600} of 0.6-0.8. All proteins used in this study were expressed at 18°C. Once at the correct density, cells were induced with IPTG to a final concentration of 1 mM and supplemented with 50% of the starting volume of antibiotic. For cells grown in TB media, 8mL of a 50% glycerol stock was added at the time of induction as well. The internal incubator temperature was typically reduced to 18°C after induction, and the protein expressed overnight (or for at least 18 hours).

2.2.3 Preparation of Selenomethionine-labelled Sgt1 1-150

Selenomethionine-labelled Sgt1 1-150 protein was prepared to obtain a heavy atom derivative dataset for SAD phasing of Sgt1:Skp1 complex crystals. The protocol used was modified from the original instruction by Van Duyne *et al*., 1993. Briefly, a 10 mL 'day culture' of T7 Express LysY/I^Q cells transformed with 'Sgt1 1-150-Untagged-pRSETa' (Table 2.5) was prepared in LB media supplemented with ampicillin to a final concentration of 100 µg/mL, and started on the morning of the day prior to large scale growth. Following at least 8 hours of growth, 1 mL of this day culture was used to inoculate 300 mL of pre-autoclaved minimal media prepared using the SelenoMethionine Medium Complete kit (Molecular Dimensions) according to the manufacturer's instructions, supplemented with ampicillin to a final concentration of 100 µg/mL and grown overnight at 37°C. This overnight culture was then used to inoculate 1 L volumes of minimal media stored in 2.5 L baffled culture flasks prepared a day before large-scale growth and pre-heated at 37°C overnight with mild shaking. Each litre was inoculated with 10 mL of the overnight culture and supplemented with antibiotic, then transferred back to a 37°C shaking incubator. Cell growth was monitored by optical density measurements at 600 nm.

Once the cells had reached an OD_{600} of 0.6-0.8, cells were supplemented with 0.5 g of an amino acid mix containing 1.0 g each of lysine, threonine and phenylalanine, 0.5 g each of leucine, isoleucine and valine, and 0.5 g of L-selenomethionine (ACROS Organics), and then returned to the incubator for 15 minutes, while allowing the internal temperature to begin cooling to 18°C for expression. Following this 15 minute incubation, expression was induced with IPTG to a final concentration of 1 mM and supplemented with 50% of the starting antibiotic concentration. Expression was conducted overnight (or at least 18 hours), and the cells lysed as described in Section 2.3.

2.3 Recombinant Protein Purification Protocols

Cell cultures prepared for large-scale expression were centrifuged at 5000 RPM using a JS-4.2 swinging-bucket rotor (Beckman Coulter). Cell pellets were resuspended by pipetting in a lysis buffer specific to the purification target (Table 2.7), using 10 mL of buffer per litre of cell culture. Benzamidine-HCl and $MgCl_2$ were added to the resuspended pellet to an approximate final concentration of 5 mM each. 50 µg aliquot of DNase I (Sigma) was then added to the cell suspension prior to cell disruption, and the complete 'lysis mix' stored on ice. Cells were disrupted using an EmulsiFlex-C5 (30000 psi/2000 bar max) (AVESTIN), with the system pressure set between 10000-15000 psi or 750-1000 bar. Homogenisation was repeated at least three times and a 10 µL sample of the lysate was retained for analysis by SDS-PAGE (Section 2.4). Lysate clarification was performed by high-speed centrifugation at 18000 RPM using a F21-8x50y fixed-angle rotor (Sorvall) for 60 minutes at 4°C and the soluble fraction stored on ice until further processing. 10 µL were also retained for analysis by SDS-PAGE.

Each chromatography column described in the subsequent sections was used according to the manufacturer's specifications. Unless otherwise indicated, all columns were attached to an ÄKTA automated chromatography system for purification using the UNICORN control software. A summary of all purification protocols followed throughout this study can be found in Table 2.6.

Protein	Construct Name (Resi.)	Molar Ext. Coeff. (M^{-1} cm^{-1})	Tag	Theo. MW (Da)	Theo. pI	Purification steps			
						1	2	3	4
Sgt1	1-121 (1 – 121)	17420	None	13970	6.83	*None attempted* – protein insoluble as shown by expression study			
	1-136 (1 – 150)	22920	None	15664	5.96	IEX (10 mL HiTrap SP FF)	SEC (HiLoad 16/60 S75)	IEX (5 mL HiTrap SP FF)	SEC
	1-150 (1 – 150)	22920	None	17462	8.77				
	1-178 (1 – 178)	22920	None	20477	8.43				
	ΔSGS (1 – 279)	39880	None	32110	8.58	IEX (10 mL HiTrap SP FF)	SEC (HiLoad 16/60 S200)	IEX (5 mL HiTrap SP FF)	SEC
		45380	C-term. Strep (SII)	33207	8.57	SII-Affinity Pulldown (5 mL StrepTactin Resin)	IEX (1 mL Resource S)	SEC (HiLoad 16/60 S200)	-
	Full length (1 – 395)	59360	None	44860	6.08	IEX (10 mL HiTrap SP FF)	SEC (HiLoad 16/60 S200)	IEX (5 mL HiTrap SP FF)	SEC
		64860	C-term. Strep (SII)	45886	6.15	SII-Affinity Pulldown (5 mL HiTrap Streptavidin HP)	IEX (1 mL Resource S)	SEC (HiLoad 16/60 S200)	-
Sgt Mutants	FL WT (1-395)	59360	N-term. His_6	47023	6.56	His_6-Affinity Pulldown (1 mL HisTrap FF)	IEX (1 mL HiTrap SP FF)	SEC (HiLoad 16/60 S200)	-
	FL L48R (1-395)	59360		47066	6.72				
	FL W58A (1-395)	53860		46908	6.56				
	FL H59A (1-395)	59360		46957	6.52				
	FL D61R (1-395)	59360		47064	6.94				
	FL R93A (1-395)	59360		46938	6.42				
	FL L126A (1-395)	59360		46981	6.56				
	FL W127A (1-395)	53860		46908	6.56				
Skp1	ΔLoop BTB/POZ (1 – 158 (Δ34 – 74))	13980	None	13230	4.76	IEX (10 mL HiTrap Q FF)	SEC (HiLoad 16/60 S75)	IEX (5 mL HiTrap Q FF)	SEC
	Full length (1 – 195)	20970	None	22330	4.35	IEX (10 mL HiTrap Q FF)	SEC (HiLoad 16/60 S75)	IEX (1 mL Resource Q)	SEC

Table 2.6: **Summary Table of Purification Steps for Recombinant Protein Purifications** – Theoretical molar extinction coefficients were derived computationally from the primary sequences of the listed constructs using the ProtParam web tool (Wilkins *et al*, 1999).

2.3.1 Affinity Purification using Streptactin (Strep/SII) Affinity Tags

Strep II-tagged proteins were partially purified from crude cell lysates as described at the start of Section 2.3. Two general methods were used for isolating the Strep-tagged protein from the cellular milieu: batch column-gravity flow and ÄKTA-controlled automated chromatography.

For batch column purification, between 2-10 mL of a 50% slurry of Strep-Tactin® Sepharose® (IBA GmbH) resin was dispensed into a 5 cm diameter Glass Econo-

Column batch chromatography column (Bio-Rad) and washed with 3-4 successive washes of 5 CV 'Strep A' (Table 2.7). Following equilibration, the soluble lysate was incubated with the resin at 4°C for at least one hour with rolling.

After incubation, the unbound fraction was collected and a 10 µL aliquot of this sample was retained for analysis by SDS-PAGE (Section 2.4). Washes of the resin were performed 5-10 times with the sequential addition of 1-2 CV of 'Strep A' and each wash fraction retained. After washing, the Strep-tagged target protein was eluted from the resin using 5-10 applications of 2-3 CV 'Strep B' (Table 2.7). All fractions were typically analysed by SDS-PAGE (Section 2.4).

ÄKTA-based purification of Strep-tagged recombinant gene products utilised a 5 mL StrepTrap HP column (GE Healthcare) for isolation of the target. The column was regenerated or equilibrated for use by washing with 4 CV of 0.5 M NaOH or HABA StrepTactin regeneration buffer R (IBA), followed by 4 CV of 'Strep A'. The cell lysate was loaded at approximate flow rates of 0.5-1.0 mL/min. The column was washed at a flow rate of 1.0 mL/min until the A_{280} reading had stabilised and the target protein then eluted over 6 CV of 'Strep B' in 2 mL fractions at a flow rate of 2.0 mL/min.

After each use, the column was stored in HABA StrepTactin buffer R (IBA) as per the manufacturer's instructions.

2.3.2 Affinity Purification using Polyhistidine (His_6) Tags

HisTrap FF/HP columns (GE Healthcare) were used for purification of His_6-tagged recombinant proteins. 1 and 5 mL columns were equilibrated in 2-3 CV of 'His-Tris A' or 'His-Complex A' buffer before use. Columns were stored in 20% ethanol when not in use. All purifications were conducted at 4°C.

Equilibrated columns were loaded with crude cell lysate via a peristaltic pump at flow rates between 0.5 and 1.0 mL/min and the unbound fraction collected.

Loading of protein that had already been subjected to some prior purification steps (e.g. purification of the Sgt1 1-150:Skp1 BTB/POZ complex for crystallisation) was conducted using a 50 mL loading loop (GE Healthcare). Crude samples were eluted with a step-elution to 50% 'His-Tris B' or 'His-Complex B' following 10-15 CV of washes in 'His-Tris A' or 'His-Complex B'. 2 mL fractions were collected for analysis by SDS-PAGE before pooling.

Buffer Type	Purification	Buffer Name	Buffer	Final pH	Salt	Reducing Agent	Further additives
Lysis/Affinity Chromatography	Untagged Sgt1 Constructs	Sgt1 Lysis	HEPES	7.0	25 mM NaCl	2 mM DTT	1 mM EDTA
	Untagged Skp1 Constructs	Skp1 Lysis	Tris-HCl	8.0	25 mM NaCl	4 mM DTT	1 mM EDTA, 10% Glycerol
	Strep II-tagged Sgt1 Constructs	Strep A	HEPES	7.0	300 mM NaCl	2 mM DTT	1 mM EDTA, 10% (w/v) glycerol
		Strep B	HEPES	7.0	50 mM NaCl	2 mM DTT	1 mM EDTA, 3 mM desthiobiotin
	His-tagged Sgt1 Constructs	His-Tris A	Tris-HCl	7.5	300 mM NaCl	-	10% (w/v) glycerol
		His-Tris B	Tris-HCl	7.5	150 mM NaCl	-	10% (w/v) glycerol, 500 mM Imidazole
	His-Complex A	His-Complex A	Tris-HCl	7.5	100 mM NaCl	-	-
	His-Complex B	His-Complex B					500 mM Imidazole
Ion Exchange (IEX)	All Sgt1 Constructs	IEX S A	HEPES	7.0	-	2 mM DTT	1 mM EDTA
		IEX S B	HEPES	7.0	1 M NaCl	2 mM DTT	1 mM EDTA
	All Skp1 Constructs	IEX Q A	Tris-HCl	7.5	-	2 mM DTT	1 mM EDTA
		IEX Q B	Tris-HCl	7.5	1 M NaCl	2 mM DTT	1 mM EDTA
Size Exclusion Chromatography (SEC)	All Skp1 Constructs	SEC Skp1	Tris-HCl	7.5	100 mM NaCl	2 mM DTT	1 mM EDTA
	All	SEC All	HEPES	7.0	150 mM NaCl	2 mM DTT	1 mM EDTA

Table 2.7: **Summary Table of Standard Buffer Solutions for Recombinant Protein Purifications**

Routine cleaning and regeneration of HisTrap columns was done using 5-10 CV washes of buffer containing 100 mM sodium phosphate buffer pH 7.4, 500 mM NaCl, 50 mM EDTA, followed by 5-10 CV of Milli-Q water (Millipore) and regeneration with 2 CV of 0.1 M $NiSO_4$.

2.3.3 ION EXCHANGE CHROMATOGRAPHY (IEX)

Ion exchange columns (HiTrap SP FF/HP, HiTrap Q FF/HP and Resource Q/S columns (GE Healthcare)) were pre-equilibrated prior to use by washing with at least 2 CV of IEX S A or IEX Q A (depending on the protein being purified) at a flow rate between 0.5 – 5 mL/min, depending on the system back pressure. This was followed by at least 2 CV of IEX S B or IEX Q B, and a final wash of twice the number of CVs of IEX S/Q A as IEX S/Q B to return the conductivity reading to a baseline for the buffer used.

The salt concentration of solvent buffers was typically diluted to below 6-7 mS/cm to facilitate binding. Samples were loaded at a flow rate of 0.5 mL/min. After loading, the column was washed with 5 CV of IEX S/Q A. Elution of bound proteins was conducted over a gradient of 0-50% IEX S/Q B over at least 10 CV, depending on the stage of the purification (typically, 10 CV for the crude purification of untagged proteins or 20-30 CV for polishing steps in the later stages of purification). 2 mL elution fractions were collected for analysis by SDS-PAGE (Section 2.4). All purifications were conducted at 4°C.

For storage, columns were equilibrated using this method then washed into filtered 20% ethanol.

2.3.4 SIZE EXCLUSION CHROMATOGRAPHY (SEC)

Preparative size exclusion chromatography was used as a final step in the purification of all proteins examined in this study. HiLoad 16/60 S200 and HiLoad 16/60 S75 (GE Healthcare) columns were used as per the specifications of the manufacturer. Protein samples ranging in volume from 2-5 mL were applied at 4°C at 0.5-1.0 mL/min flow rates with the appropriate molecular weight exclusion column connected. Prior to loading, the column was equilibrated by passing through 1.5 CV of 'SEC All' or 'SEC Skp1' buffer (Table 2.7) at a flow-rate of 0.8-1.0 mL/min, with the buffer chosen based on the stage of purification (all final SEC steps were conducted using 'SEC All' buffer, while

intermediate size-exclusion steps for all Skp1 constructs used 'SEC Skp1' buffer to facilitate the second ion exchange step (Table 2.7)). Proteins were eluted over 1.2 CV in the same buffer used for equilibration at a flow-rate of 0.5 mL/min, and elution fractions collected.

2.3.5 Protocols for the Preparation of Protein Complexes

For the complex of Sgt1 1-150 and Skp1Δ BTB/POZ, both proteins were expressed and purified separately as described previously, and the complex reconstituted *in vitro*. Based on biophysical and mass spectrometry experiments on the complex, Sgt1 1-150 and Skp1Δ BTB/POZ were mixed in a 2:1 molar ratio for stoichiometric complex formation. The mixture was allowed to equilibrate on a roller kept at 4°C for at least one hour before concentration and preparation for crystallisation (Section 2.3.7 and 2.6.1).

2.3.6 Purification of Selenomethionine-labelled Sgt1 1-150

Purification of selenomethionine-labelled Sgt1 1-150 followed the same protocol as indicated in Table 2.6 for unlabelled Sgt1 1-150. The procedure for concentrating the protein for crystallisation was identical to that described for the unlabelled protein (Section 2.3.7).

2.3.7 Quantitation of Protein Yield from Recombinant Expression and Sample Concentration

Following purification, the final eluent pool was quantified by absorbance at 280 nm using a NanoDrop 1000 spectrophotometer (Thermo Scientific) according to the manufacturer's instructions and the accompanying software. Typically, two different dilutions (a combination of a 1:2, 1:5 and/or 1:10) were prepared and tested, with samples measured twice and fresh applications of protein sample both times. The average value for each of these recordings was then used in the calculation of the protein concentration.

Using the average corrected absorbance reading at 280 nm, the protein concentration was then calculated using the following rearrangement of the Beer-Lambert equation, assuming a 1 mm path length:

Eq. 2.1
$$Concentration\,(\mu M) = (\frac{\varepsilon}{A_{280}}) \bullet 10^5$$

Where ε is the specific molar extinction coefficient (in M^{-1} cm^{-1}) of the protein construct and A_{280} the absorbance reading recorded by a NanoDrop spectrophotometer (Thermo Scientific). Theoretical molar extinction coefficients were calculated from the primary amino acid sequence of the protein being quantitated using the ProtParam web tool (Wilkins *et al.*, 1999).

To obtain the appropriate concentrations of protein for structural and biophysical studies, preparations of recombinant proteins were concentrated by ultrafiltration at 3500-4000 x g using 10000 Da molecular weight cut-off (MWCO) 4 or 15 mL centrifugal filter units (Amicon). All samples were kept at 4°C during centrifugation. Typically, protein samples were over-concentrated to or close to the dead-volume of the filter unit and then diluted in 'SEC All' buffer to the desired concentration in a separate vessel. The final concentration was confirmed by taking the average corrected absorbance reading for dilutions of a neat protein sample stock at 280 nm as described previously.

2.4 Analysis of Proteins by Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Pre-cast 8-12% Bis-Tris polyacrylamide gels (Invitrogen) were prepared and run as instructed in the manufacturer's guide book for SDS-PAGE analysis. For hand-cast gels, the Tetra Electrophoresis system (BioRad) was used. Reagents for 10%, 12% and 15% polyacrylamide gels were combined in volumes detailed in Table 2.8 with the exception of TEMED (Sigma-Aldrich) and 10% ammonium persulphate solutions, which were added immediately prior to pouring.

Gel Percentage (%)	**Resolving Gel**			**Stacking Gel**
	10	**12**	**15**	
Reagent	Volume of Reagent (mL)			
1.5 M Tris-HCl pH 8.8	2.5	2.5	2.5	-
1.0 M Tris-HCl pH 6.8	-	-	-	0.63
40% Polyacrylamide solution (BioRad)	3.3	4.0	5.0	0.83
10% Sodium Dodecyl Sulphate	0.1	0.1	0.1	0.05
Distilled Water	4.0	4.3	3.3	3.40
TEMED (Sigma-Aldrich)	0.008	0.008	0.008	0.008
10% Ammonium Persulphate	0.100	0.100	0.100	0.050
Approximate Total Reagent Volume (mL)	10.0			5.0

Table 2.8: **Composition of Hand-Cast Polyacrylamide Gels for SDS-PAGE**

Protein samples to be analysed by SDS-PAGE were mixed with a 3X preparation of gel-loading buffer (100 mM Tris pH 6.8, 30% glycerol, 3% SDS, 0.01% bromophenol blue, 350 mM DTT) as follows:

- For crude fractions (including whole cell, soluble and flow through fractions), a 1:20 dilution of sample was prepared using appropriate volumes of MilliQ water (Millipore) and 3X GLB.
- For partially purified protein samples from either a batch or ÄKTA-based purification, samples were mixed in a 1:1 volumetric ratio of sample to 3X GLB.

Typically, 5-15 µL were loaded into the wells of the hand-cast gels, and 12 µL into the well of pre-cast gels (Invitrogen). Electrophoretic separation using the MiniProtein system and a PowerPac™ basic power supply (BioRad), was run at 200 V typically for 50-60 minutes depending on the percentage of the resolving gel. Gels were stained with Coomassie-stain (40% ethanol, 10% acetic acid, 1% Coomassie Blue stain (dry powder) (BioRad)) and destained with buffer containing 20% ethanol and 10% acetic acid. Gels were visualised on a light box or scanned for electronic visualization.

For size approximation by SDS-PAGE, 5-10 µl of either SeeBlue® Plus2 (Invitrogen) or Page Ruler Prestained Protein Ladder (Thermo Scientific) were used as molecular weight markers.

2.5 Biophysical Analysis of Purified Recombinant Proteins

Recombinant gene products were expressed and purified as previously described (Table 2.6 and Table 2.7). For size exclusion chromatography-coupled multi-angle light scattering (SEC-MALS) analysis, proteins were dialysed into the buffer used for analysis overnight and prepared to a minimum concentration equivalent to 2.0 mg/ml. Samples for sedimentation velocity analytical ultracentrifugation (SV-AUC) were not dialysed prior to analysis, but were instead passed over a size exclusion chromatography column equilibrated in the buffer used for analysis. In all SV-AUC experiments, fresh DTT was found to be vital to maintain a flat baseline during data recording. Therefore, all samples were prepared for SV-AUC no more than 18 hours prior to analysis to minimise oxidation.

2.5.1 Sedimentation Velocity Analytical Ultracentrifugation (SV-AUC)

Hydrodynamic analysis of Sgt1 and Sgt1:Skp1 complexes was carried out by SV-AUC using a ProteomeLab XL-I (Beckman Coulter) analytical ultracentrifuge. All runs were carried out at either 42000 or 60000 RPM, employing an An-50 Ti eight-cell or An-60 Ti four-cell rotor (Beckman Coulter) respectively. Analytical cells (Beckman Coulter or Spin Analytical) were assembled as specified by the manufacturer; 12-mm centrepieces (for An-50 Ti rotors used at 42000 RPM speed) were torqued at 120 inch-pounds, and centerpieces for An-60 Ti rotors (60000 RPM) were torqued at 135 inch-pounds prior to filling the separate channels. Solvent and sample-solvent channels were filled with 330 or 420 μL of buffer and protein sample respectively, depending on the specification of the centrepieces (420 μL for 12-mm and 330 μL for 3-mm centrepieces). Finally, filled sectors were sealed with screw plugs supplied with the individual cells and balanced to within 0.5 g of each other.

For each run, the ultracentrifuge was calibrated according to the manufacturer's guidelines. Intensity scans were conducted to ensure that sufficient light was passing through each of the solvent and sample-solvent channels. Following

calibration, cells were recovered and rotated in-hand to resuspend the sample; the ultracentrifuge was then equilibrated to temperature and left under vacuum for an additional hour after reaching the analysis temperature to ensure full system equilibration. For data collection, the data collection software was set to obtain 999 scans for every cell using both interference and absorbance optical systems (with absorbance at 280 nm). All analysis was conducted at 20°C and the scans run to completion.

All data analysis was carried out in Sedfit (Schuck, 2000), which provides a means of qualitatively assessing experimental protein sedimentation data relative to defined protein association models. This is achieved by solving the Lamm equation (Equation 2.2), which describes the evolution of the concentration of species over time when under the influence of a centrifugal force, $\omega^2 r$.

Eq. 2.2

$$\frac{dc}{dt} = \frac{1}{r}\frac{d}{dr}\left[rD\frac{dc}{dr} - s\omega^2 r^2 c \right]$$

The change in solute distribution with time is dependent on the diffusion coefficient, *D*, as well as the total concentration of particles, *c*, which in turn determine the sedimentation coefficient, *s*, measured in Svedberg units (S, where 1 S is 10^{-13} seconds).

Centrifugation through a solution of a known density and viscosity results in either a concentration gradient or distinct sedimentation boundaries for particles in solution (because of association/dissociation or sample heterogeneity) which are determined by the shape and size of the analytes. This relationship is defined by the Svedberg equation for the motion of particles in solution (Equation 2.3):

Eq. 2.3

$$s = \frac{M(1 - vbar \bullet \rho)}{N \bullet f}$$

Where *M* is the molecular mass of the particle, *vbar* its partial specific volume, ρ tbe solvent density, *f* a relative shape factor (or frictional ratio) and N is Avogadro's number. This relationship states that a particle's sedimentation rate (i.e. s-value) is determined by both its mass and its shape, which are acted upon by centrifugation and diffusion, as described by the Lamm equation. All experimental sedimentation velocity data can therefore be described by satisfying the Lamm equation and the relationships ascribed by the Svedberg equation.

While it may therefore be theoretically possible to implement a brute-force fitting of the Lamm equation parameters to describe the experimental data, this approach is hampered by the physical experimental conditions (such as the shape of centriguation vessels, which results in radial dilution) and broadening of the sedimentation boundaries due to the effects of diffusion, caused directly by the counter-action of the particle against the centrifugal field.

In Sedfit, the experimental data is therefore fitted through iterative numerical simulations that take into account known information about the solvent, analyte (including the partial specific volume, v-bar, or particle density) and the physical parameters of the experimental setup, including the centrifugal force, rotor acceleration, and the size and shape of the solution column. A typical starting point is the *continuous c(s) distribution* (Equation 2.4), which describes mixtures of non-interacting molecules not exhibiting non-ideal behaviour. Sedfit attempts to minimise this function by solving for the parameters *s* and *D* within the Lamm equation, under the assumption that all species in the sample bear a similar shape (i.e. *frictional ratio*). Each s-value can therefore be given an approximate molecular weight and diffusion coefficient (since size and shape influence the diffusion of a particle).

Eq. 2.4
$$\underset{c(s)}{Min}\left\{\sum_{i,j}\left[a(r_i,t_j)-\int c(s)L(s,D(s),r_i,t_j)ds\right]^2\right\}$$

In the continuous c(s) analysis, Sedfit uses a variation of the Svedberg relationship between particle mass and sedimentation (Equation 2.5) in order to determine *D*, and uses this to satisfy the Lamm equation for every value of s. The result is represented as a series of peaks in a differential sedimentation coefficient distribution, with the height of each corresponding to the size of population in a sample. In essence, the analysis attempts to correlate solutions of the Lamm equation with all s-values in the fitting range for a given (and fixed) frictional ratio. Extensions of this model will take into account other known information about the mass or shape of the particle, in order to more accurately model the experimental sedimentation boundaries.

Eq. 2.5

$$D(s) = \frac{sRT}{M(s)(1 - vbar \bullet \rho)}$$

In practice, for each iteration of data analysis, menisci and bottom of the solution column were first manually set, and the meniscus then fitted, and manually adjusted if considered unreasonable. A standard c(s) distribution fit was applied in the first instance, and then evaluated for general fit, the presence of smaller molecular weight species that might be better described by a c(s) distribution with one discrete component, and higher molecular weight aggregates. In instances where there was clear separation of peaks in the distribution, a bimodal frictional ratio model was applied to optimise frictional ratios for both species. In analysing data, the fitting range was always adjusted to centre the peak distribution (typically, from 0.5-5.0 S or 0.5-10 S), provided no minor smaller or higher molecular weight species could be detected by the data analysis. Partial specific volume, buffer viscosity and buffer density were calculated from the amino acid sequence and buffer composition using Sednterp (Laue *et al*., 1992).

2.5.2 Size Exclusion Chromatography-coupled Multi-angle Light Scattering (SEC-MALS)

SEC-MALS is a means of determining the absolute molar mass of a particle in solution, given the difference between the refractive index of the solute particle and the pure solvent. Light scattering caused by a particle in solution is measured at multiple angles by eighteen detectors placed around the sample cell. As the incident light, I_0, is scattered by the particle, constructive and destructive interference results in differences in the intensity of scattered light, I_S, at the various detector angles. In addition to having an angular-dependence, the intensity of the scattered light is also directly proportional to the weight-average molecular weight, M, and concentration of the particle, C_B. This variation in the light intensity is expressed in the Rayleigh ratio, R_θ, which has an additional dependence upon the laser wavelength, λ, and the refractive index of the pure solvent, n_0, measured in the change in the refractive index of the particle solution with concentration (i.e. $\delta n/\delta C_B$).

Eq. 2.6

$$R_\theta = \left[\frac{4\pi^2 n_0^2 \left(\frac{\delta n}{\delta C_B} \right)^2}{\lambda^4 N_A} \right] C_B M$$

As measurements of absolute concentration or mass are expressed in g/mL or g/mol, respectively, the Rayleigh equation additionally factors in Avogadro's number, N_A. The Rayleigh ratio is therefore expressed as the ratio of scattered light to incident light intensity per unit angle, scattering volume and incident light intensity. However, the Rayleigh ratio in itself does not determine the scattering intensity of the particle (or particles) in solution, and instead is used to measure, for example, the intensity of light scattered by the pure solvent relative to the intensity of the incident beam (referred to as R_{sol}, which is the Rayleigh ratio for light scattering of the pure solvent). In effect, particle light scattering is directly proportional to the excess scattering caused by the particle

over the pure solvent. The true scattering intensity of the particle is therefore determined by the excess Rayleigh ratio, R_{ex}:

Eq. 2.7
$$R_{ex(\theta,C_B)} = wMK\left[1-\frac{(qR_g)^2}{3}\right](1-2B_2w)$$

Where both *w* and *K* are constants specific to the sample and experiment (i.e. the laser wavelength, sample concentration and particle mass). Since R_{ex} is the difference between the total scattered light intensity and the scattered light intensity from the pure solvent at any given angle (i.e. $R_{ex} = R_\theta - R_{sol}$), it is possible to get estimates for the radius of gyration (a shape factor), R_g, second virial coefficient, B_2, and molar mass, by performing a Zimm analysis on light scattering data gathered from a series of concentrations. Alternatively, a partial Zimm analysis can be implemented on light scattering data from a single concentration. This type of analysis relies on strong scattering intensity at multiple angles, while the accuracy of these calculations is further contingent on monodispersity within the sample.

SEC-MALS experiments were carried out at the Institute of Cancer Research, London on a set up comprising a Superdex 75 10/300 GL (GE Healthcare) gel filtration column coupled to a DAWN® HELEOS® II (Wyatt Technology) light scattering instrument and Optilab T-rEX Refractive Index Detector (Wyatt Technology), with a laser wavelength of 658 nm. All runs were carried out at room temperature and samples injected under a continuous flow of 0.5 mL/min. Data analysis was conducted using ASTRA 5.3.4 software (Wyatt Technology), employing the in-built automated tool to conduct a partial Zimm analysis, using the Zimm formulation of the Rayleigh-Debye-Gans approximation for the absolute molar mass of polymers. In order to minimise the error of the molecular mass determination, signal baselines were corrected manually for every detector (i.e. every scattering angle) and the dataset restricted to detector angles that gave a strong intensity reading. Poor signals were manually ommited from the analysis.

2.5.3 Analytical Size Exclusion Chromatography

SEC for analytical purposes used a Superdex 200 10/300 GL size exclusion column (GE Healthcare) connected to an ÄKTA Purifier 100 at flow-rates of 0.5 mL/min. Columns were typically pre-equilibrated prior to analysis in 'SEC All' buffer. Size approximations of analytes were obtained from a calibration curve generated by a Gel Filtration HMW Calibration Kit (GE Healthcare) as indicated in the accompanying manufacturer's guidelines.

2.5.4 Native Electrospray Ionisation Mass Spectrometry

Native mass spectrometry experiments were carried out by Richard Kerr (University College London) using a Waters Synapt HDMS equipped with a 32k quadrupole generator. Instrument calibration was carried out using 24 mg/mL CsI diluted in 30% isopropanol. Prior to analysis, samples of purified Sgt1 or Skp1 constructs were concentrated using Amicon Ultra 1.5 mL 10 kDa MWCO centrifugal concentrator columns (Millipore) and finally buffer exchanged into 150 mM ammonium acetate pH 7.0 via Micro Bio-Spin chromatography columns (Biorad). Samples were quantified using a Qubit 2.0 fluorometer (Invitrogen). Final concentrations were subsequently prepared by dilution with 150 mM ammonium acetate (pH 7) solution.

Parameter	**Setting**
Capillary voltage	0.95-1.25 kV
Sample cone	80 V
Trap collision voltage	5 V
Transfer collision voltage	10 V
Backing pressure	5.75 e^{0} mBar
Time-of-flight analyser pressure	2.37 e^{-5} mBar
Acquisition time	2 minutes, 2 second scans

Table 2.9: **Experimental Parameters for Native Nano ESI-MS Experiments** – Mass spectrometry experiments were designed and carried out by Richard Kerr (University College London).

For analysis, a total of 2.5 µL of sample was loaded into in-house prepared gold coated capillaries and ionised using a nano ESI source. In order to standardise

analyses and preserve non-covalent interactions, MS parameters were set as in Table 2.9.

For analysis of Sgt1:Skp1 complexes, separate solutions of Sgt1 and Skp1 proteins were mixed in a 2:1 ratio following quantitation and incubated on ice for 60 minutes. Parameters were maintained as in Table 2.X, except that the isolated +21 species (at 5521.68 m/z) was subject to a collision voltage ramp from 10 V to 140 V in 10 V increments and data acquisition was performed for 1 minute with 2 second scans.

2.6 Protein X-Ray Crystallography

All targets for crystallisation were analysed by SDS-PAGE for purity and, as for protein complexes, to confirm the putative ratio of components generated by *in vitro* reconstitution.

2.6.1 Preparation of Crystallisation Targets

Protein concentrations equivalent to 16 and 32 mg/ml were trialed for crystal formation of the Sgt1:Skp1 complex at both 4°C and 16°C. Sgt1 1-150 and Sgt1 1-136 was crystallised at concentrations between 30-60 mg/ml at 16°C. These final concentrations were prepared as previously described in Section 2.3.7.

2.6.2 Preparation of Crystallisation Trays and Monitoring Procedure

MRC 2 96-well format crystallisation plates (Hampton Research) were used for broad screening of the conditions provided by Jena Bioscience supplied JCSG, PACT, Classic I HTX, Classic II HTX and pH screens. Reservoir solutions were dispensed into the deep-wells of MRC 2 plates using a Liquidator 96 (Rainin) liquid dispenser to a reservoir volume of 80 µL and 150 nL total volume drops prepared in the two subwells using a Mosquito® dispensing robot (TTP Labtech). The Mosquito robot was programmed to dispense 1:1 (75 nL:75 nL) and 2:1 (100 nL:50 nL) 'protein:reservoir solution' drops in the top and bottom sub-

wells respectively. Plates were sealed and stored in temperature controlled rooms maintained at 16°C or 4°C.

Plates stored at 16°C were placed into and monitored by a Desktop Minstrel (Rigaku) plate imaging system with dual UV and visible light images recorded over a one month period. Crystallisation plates stored at 4°C were monitored manually via a light microscope; if crystals were found, the plates were placed into the Minstrel to determine, via the UV imaging system, whether or not these were protein crystals. Harvesting of crystals is described in Section 2.6.3.

Optimisation of crystallisation conditions was carried out either in MRC 2 96-well (Hampton Research) or MRC Maxi 48-well format plates (Jena Bioscience). Optimisation conditions were determined automatically by the optimisation component of the Rigaku plate imaging software. Reservoir solutions were prepared manually or via a Multiprobe II Plus HT EX (Perkin Elmer) liquid handling robot. All optimisation screens were kept at 4°C until use.

2.6.3 Freezing Protocol for Macromolecular Crystals

Cryoprotectant solutions were screened initially, but ethylene glycol was found to be the most effective and therefore used in freezing all crystals obtained throughout this study. For crystals harvested from commercial screens (JCSG, PACT, Classics or pH screen (Jena Biosciences)), 0.5 µL of reservoir solution was added directly to the crystallisation drop to aid in the recovery of crystals for freezing using appropriately sized nylon or grid loops.

Otherwise, a cryoprotectant solution was made up using the components of the reservoir solution plus 20% (w/v) ethylene glycol and a 1 µL drop placed onto a 22 mm circular cover slip. Crystals were taken from their drops and soaked in the cryoprotectant for 10-30 seconds by depositing the crystal into the solution, then rapidly recovered into a loop and flash frozen in liquid nitrogen. Crystals were stored in liquid nitrogen until transport to a synchrotron X-ray source.

2.6.4 COLLECTION AND PROCESSING OF MACROMOLECULAR DIFFRACTION DATA

Native, and single- and multiple-wavelength anomalous dispersion (SAD/MAD) datasets were collected at the Diamond Light Source synchrotron. A native dataset of Sgt1:Skp1 complex crystals was collected on the IO4-1 fixed wavelength beamline. Dataset collection was determined using an EDNA-derived strategy to maximise resolution and dataset redundancy (Incardona *et al.*, 2009) (Table 2.10).

Wavelength (Å)	0.9163	**Max. Resolution** (Å)	3.26
Omega Start (°)	143	**Omega Oscillation** (°)	0.20
Time Per Image (sec)	0.08	**Transmission** (%)	100
Number Of Images	595	**Overlap**	0

Table 2.10: **Native Dataset Collection Strategy for Crystals of an Sgt1:Skp1 Complex**

Derivative datasets for Sgt1:Skp1 complex crystals were collected on the IO4 tuneable beamline. A fluorescence scan was first conducted to determine the optimum wavelength for collecting a dataset at the selenium absorption edge. Data was collected using the following EDNA-derived strategy to maximise the quality of the dataset and its multiplicity:

Wavelength (Å)	0.979494	**Max. Resolution** (Å)	2.6
Omega Start (°)	0	**Omega Oscillation** (°)	1.2
Time Per Image (sec)	0.932	**Transmission** (%)	71.373999
Number Of Images	301	**Overlap**	0

Table 2.11: **SAD Dataset Collection Strategy for Crystals of an Sgt1:Skp1 Complex**

A native dataset for crystals of Sgt1 1-136 was collected on the IO4-1 monochromatic beamline. To maximise the resolution, the beam transmission was set at 100% and a line scan conducted across the crystal to obtain a complete dataset (Table 2.12).

Wavelength (Å)	0.9163	**Max. Resolution** (Å)	3.995743
Omega Start (°)	45	**Omega Oscillation** (°)	2
Time Per Image (sec)	2.5	**Transmission** (%)	100
Number Of Images	60	**Overlap**	0
Line Scan Parameters	Line scan from (-831.10, 144.86, 150.34) to (-734.10, 146.85, 146.30) in 12 wedges of 5 images per wedge.		

Table 2.12: **Line Scan Data Collection Strategy for Crystals of Sgt1 1-136**

Dataset processing and reduction was performed locally at the Diamond Light Source via xia2 (Winter, 2010).

2.6.5 PHASING STRATEGIES

For obtaining starting phases from anomalous datasets, the product *mtz* file from xia2 processing was fed into the Auto-Rickshaw pipeline (Panjikar *et al*., 2005) to obtain selenium atom locations, starting phases and an initial model for manual rebuilding in Coot (Emsley *et al*., 2010).

Molecular replacement for the dataset of the Sgt1 TPR domain on its own was conducted using Phaser (McCoy *et al*., 2007) in the CCP4 GUI (Winn *et al*., 2011), using a partial model of the TPR domain (previously solved by SAD phasing and the Auto-Rickshaw pipeline (Panjikar *et al*., 2005)). The successful MR search was performed using a truncation of the Sgt1 TPR domain to residue 121, with three copies of the monomer found in the asymmetric unit. The rotation and translation search cut-off was set at 85%, and the solvent content set to 0.60 (60%) based on the Matthews coefficient (Kantardjieff *et al*., 2003). Resolution cut-offs were not altered.

2.6.6 STRUCTURE SOLUTION AND REFINEMENT

Qualitative assessment of phased maps and partial models was performed in Coot or WinCoot (Emsley *et al*., 2010).

Refinement of partially complete models derived through either SAD phasing in the Auto-Rickshaw pipeline (Panjikar *et al*., 2005) or MR in Phaser (McCoy *et al*., 2007) (Section 2.6.5), was performed in BUSTER (Bricogne, 1993). Each round of refinement was followed by manual building and improvement of the model in Coot (Emsley *et al*., 2010), including the correction of side-chain placements and bulk sections of misplaced secondary structure.

Non-prerequisite BUSTER parameters were used as standard to aid refinement of the Sgt1:Skp1 complex model, including:

- **-autoncs:** Analyses sets of non-bonded atoms that are within 5.5 Å of each other, and evaluates the inter-atomic distances of equivalent sets between chains. If the distances are similar between chains, the function specifies non-crystallographic symmetry (NCS) for that region and applies local structural similarity restraints (LSSR).
- **'-L':** Performs a search in the map for regions of continuous density and builds water molecules into these regions as placeholders for "areas of interest".
- **'-M TLSbasic'** and **'–TLS':** Separates a macromolecular model into multiple rigid bodies that are undergoing vibrational motion (TLS, translation/libration/screw motion), effectively specifying different functional groups for individual refinement. '-M TLSbasic' is used to first obtain the TLS groups, and '-TLS' used for subsequent rounds of refinement. Only once the R_{Work} of the Sgt1:Skp1 structure had stabilised around 30%, were TLS parameters was used.
- **'-Gelly protgeo_option_chiralrestraint_from_equilib.dat'**: This function improves the geometry of the input model and nullifies the effect of Cβ deviations on the Molprobity geometry score (Chen *et al*., 2010). The script was obtained from the GlobalPhasing Buster Wiki (http://www.globalphasing.com/buster/wiki/index.cgi?BusterFAQ).

In addition, the following parameters were used in the refinement of the Sgt1 TPR domain structure:

- **'-target'**: Applies restraints onto chains in the unknown structure to regions with similar inter-atomic distances in a specified target structure (similar to '-autoncs' but using an external input structure).
- **'-RB'**: Specifies rounds of rigid body refinement to occur before the macro- and micro-cycles of restrained refinement.

Water molecules were added only to the 2.8 Å structure of the Sgt1:Skp1 complex, and were placed automatically using the '-WAT' command in BUSTER. Placement of these was then altered manually in Coot to reflect the electron density of the map. Geometry refinement and final checks were carried out using the Molprobity web tool (Chen VB *et al*., 2010; Davis *et al*., 2007).

2.7 Analysis of Protein:Protein Interactions by Pulldown

Pulldown assays were performed using His_6-tagged Sgt1 (wild type (WT) and mutants) and full length wild type Skp1 protein prepared as outlined in Section 2.3, with the aim of probing the interface of the Sgt1:Skp1 structure and the contribution of Sgt1 dimerisation to this interaction. 120 µL of Ni Sepharose High Performance IMAC resin (GE Healthcare) was aliquoted into 1.5 mL tubes (Eppendorff) per pulldown reaction. 1 mL of 'PD washing buffer' (25 mM HEPES pH 7.0, 100 mM NaCl) was added to each aliquot and the tubes vortexed for 10 seconds, before pelleting the resin by centrifugation at 3000 x g for 3 minutes and the supernatant aspirated using a gel-loading pipette tip (to minimise resin loss with each wash step). This wash step was repeated twice. Subsequently, 200 µL of 'PD blocking buffer' ('PD washing buffer' with 1.0 mg/mL final concentration of BSA) was added to each tube and incubated on a roller at 4°C for one hour, after which the resin was once again pelleted by centrifugation, and the supernatant aspirated and discarded.

His_6-tagged Sgt1 stocks stored in 'SEC All' buffer were prepared separately to a final concentration of 45 µM using 'PD washing buffer'. 200 µL of each stock was added to separate tubes and 200 µL of 'PD washing buffer' to an additional tube for the control reaction. All separate tubes were returned to a roller at 4°C and incubated for one hour to saturate the nickel resin prior to adding Skp1.

Full length Skp1 protein was prepared at a concentration of 405 µM using 'PD washing buffer' that had been supplemented with 2 mM DTT and 30 mM imidazole. After allowing the His_6-tagged Sgt1 proteins to incubate for one hour, 50 µL of full length Skp1 was added to each reaction tube (including the control)

to probe the Sgt1:Skp1 interaction. This pulldown mix was incubated for a further hour before processing samples for analysis by SDS-PAGE. 10 µL of the reaction mix ('load') were retained for analysis by SDS-PAGE.

Reaction mixes were centrifuged at 3000 x g for 3 minutes in a temperature controlled bench-top centrifuge set at 4°C. The supernatants were retained as the 'unbound' fraction for analysis by SDS-PAGE. Three wash steps were performed as described previously using 500 µL of 'PD washing buffer' for each wash step. Proteins were eluted by the addition of 50 µL of 3X GLB and incubating each reaction tube at 95°C for 3 minutes ('bound' sample). 'Load' and 'bound' samples were analysed in an appropriate order by SDS-PAGE.

CHAPTER 3: BIOPHYSICAL CHARACTERISATION OF SGT1 AND THE SGT1:SKP1 INTERACTION

Through its modular architecture, Sgt1 performs a unique scaffolding function in the context of kinetochore assembly, bridging the CBF3 components Skp1 and Ctf13, with Hsp90, and initiating the CBF3 complex for DNA binding. Structural information for Sgt1 is scarce, existing solely for its CS domain from solution NMR (Lee *et al*., 2004) and in crystal structures in complex with the N-terminal domain of Hsp90 (Zhang *et al*., 2008; Zhang *et al*., 2009). While this has shed light on the structural basis for the chaperoning role of the Hsp90-Sgt1 pair (including a preference for the ADP-bound, lid open conformation of Hsp90, also shown by Catlett *et al*., 2006), it imparts little on the structural basis for Sgt1's behaviour in solution. It has been shown that Sgt1 has an elongated shape and forms dimers, which is critical to Sgt1's function *in vivo* (Bansal *et al.*, 2008). The extent of self-association differs between species, with the yeast orthologue exhibiting a similar behaviour *in vitro* as the barley and cress proteins, while human Sgt1 shows no oligomerisation behaviour *in vitro* (Nyarko *et al.*, 2007). In plants, this self-association is mediated by an intra-molecular disulphide bond that forms between helix 3A and the capping helix, and prevents oligomerisation (Nyarko *et* al., 2007). One of the aims of this project was to further deconstruct the solution behaviour of Sgt1, through the use of numerous biophysical techniques, including analytical ultracentrifugation (AUC) and size exclusion chromatography-coupled multi-angle light scattering (SEC-MALS).

3.1 PRIMARY SEQUENCE ANALYSIS OF *SACCHAROMYCES CEREVISIAE* SGT1

Previous experiments have shown that dimerisation of Sgt1 is required for yeast viability and kinetochore formation (Bansal *et al.*, 2009). As part of the method for identifying the domain responsible for self-association and Skp1 binding, it was determined that all three putative TPR repeats in the N-terminal region were required for forming homodimers and complexes with Skp1 (one of the structural components of the CBF3 inner kinetochore complex in yeast). In the

absence of structural information on the TPR domain of Sgt1, an initial assessment of the primary sequence and secondary structure of the N-terminus was undertaken, with the aim of identifying features typical of oligomerising TPRs.

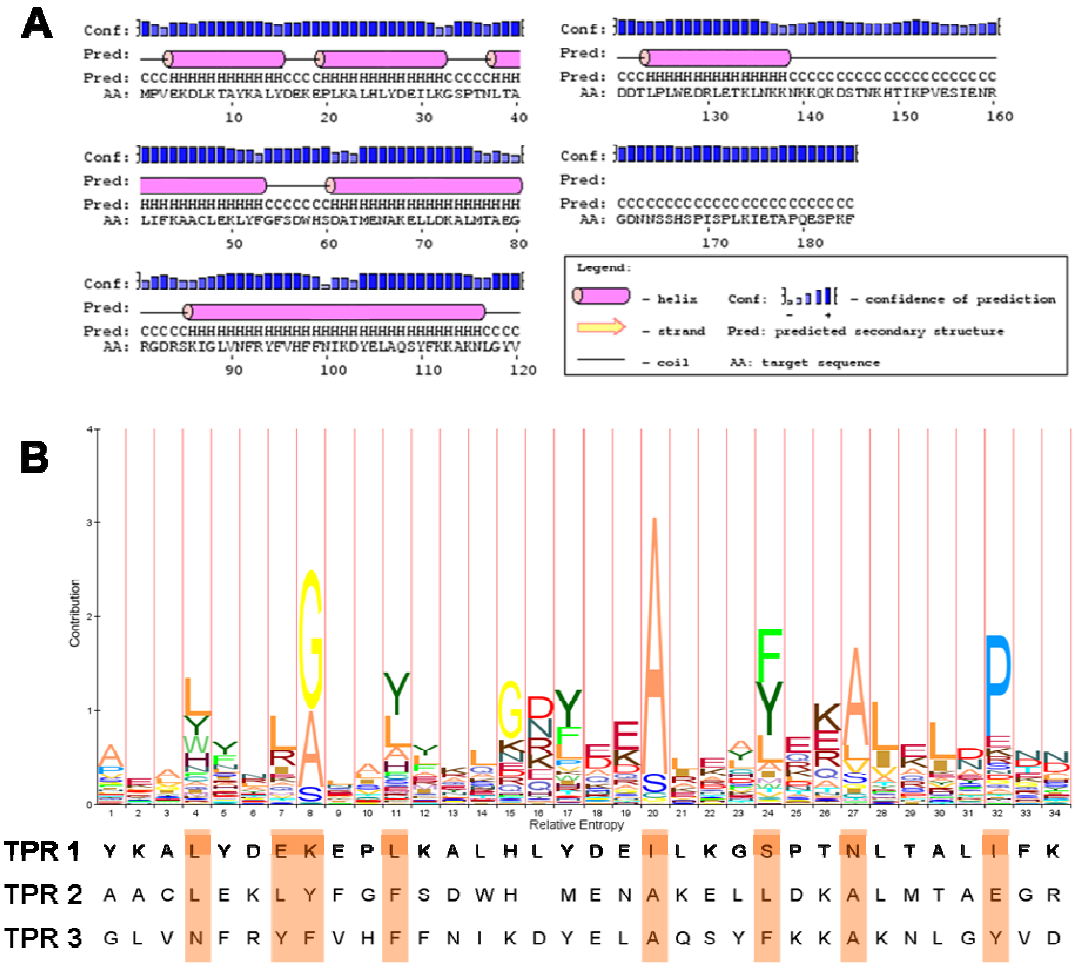


Figure 3.1: **Sequence Analysis of Sgt1** – **(A)** The N-terminus of Sgt1, up to residue 178, is predominantly alpha helical as evidenced by this secondary structure prediction using the program PSIPRED (Buchan *et al.*, 2010). **(B)** The HMM Logo of TPR domains (Pfam family PF07719) (Punta *et al.*, 2012) highlighting the preferred residue at each position in a single TPR repeat (TPR_2, Pfam family PF07719). The putative TPR repeats of Sgt1 are aligned against the HMM logo underneath, with consensus positions highlighted in **Orange**, based on D'Andrea *et al*, 2003).

There are a few known features that contribute to TPR self-association. These include residues in the loop regions of a TPR domain, exemplified in the tyrosine-mediated Pi-stacking interactions found in the inter-TPR loops of the engineered consensus three-repeat TPR structure, CTPR3Y3 (PDB ID: 2WQH; Krachler *et al.*, 2010) (a triple Tyrosine mutant of the consensus TPR, CTPR3 (PDB ID: 1NA0;

Main *et al.*, 2003)) or the hydrogen bond-network of residues in the intra-TPR regions of yeast Bub1 (PDB ID: 3ESL; Bolanos-Garcia *et al.*, 2009). This type of interaction leads to a stacking of TPR protomers, and an extended concave binding surface. A second type of interface involves parallel packing of helices of adjacent protomers, which results in the extension of the superhelical twist of the TPR. This is seen in the structure of CTPR3Y3, where displacement of the capping helix leads to packing of 3B helices in an end-on C-terminal association, and also in the type III secretion chaperone SycD (Büttner *et al.*, 2008). In the latter, this type of interface is mediated in one of two ways: either by head-to-head N-terminal packing of TPR 1 (PDB ID: 2VGY) or head-to-tail packing of TPRs 1 and 3 (PDB ID: 2VGX). Lastly, the structure of the six-repeat TPR protein PilF, involved in pilus biogenesis, shows an association of convex faces resulting in a back-to-back conformation of two adjacent TPR domains (Kim *et al.*, 2006). The sequence of Sgt1 was therefore investigated for evidence of these interfaces in the Sgt1 homodimer.

A secondary structure prediction of Sgt1 (Figure 3.1, A) using PSIPRED (Buchan *et al.*, 2010) shows that the domain is largely helical, with six predicted helices. This secondary structure is confirmed on the basis of a comparison with the Pfam HMM logo (Figure 3.1, B) (Punta *et al.*, 2012) which predicts three TPR repeats (helix-turn-helix motifs). This comparison suggests that the assignment of the individual TPR helices and loop regions in the PSIPRED prediction is incorrect. PSIPRED also predicts an additional helix after the three TPR repeats inferred by the HMM logo. Aligning Sgt1 against CTPR3 indicates that this is equivalent to a capping helix (Figure 3.2, A). Sequences C-terminal to residue 139 are predicted to be unstructured (Figure 3.1, A).

In CTPR3, the capping helix is packed against the last helix in the TPR core. This is mediated by a critical hydrogen bond between Asp107 in the last helix of the core TPR and Asn109 in the capping helix, which maintains tight packing between the two helices by stabilising the conjoining loop region (Main *et al.*, 2003) (see Section 1.2.1.2 for further details on the structure of the TPR domain of CTPR3).

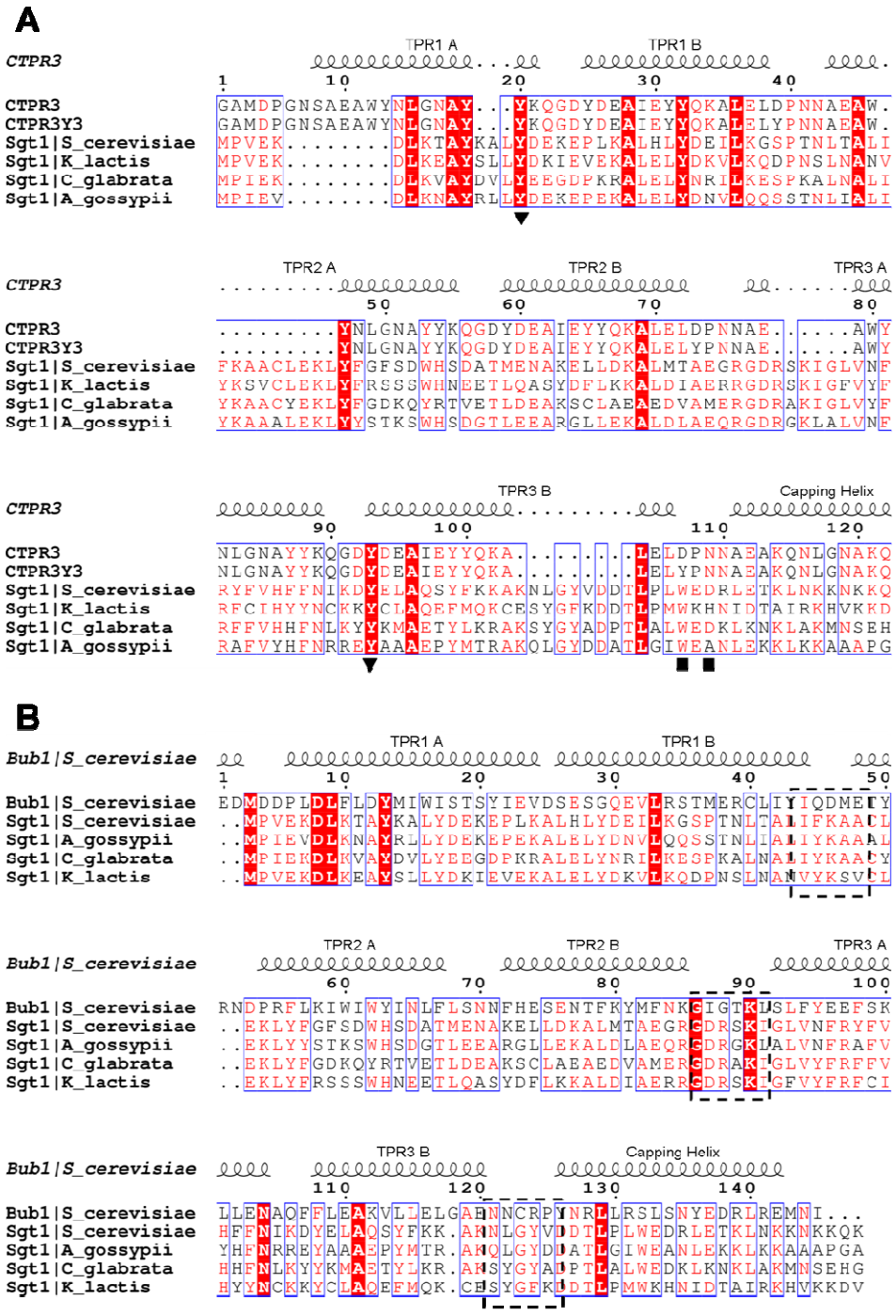


Figure 3.2: **Primary Sequence Alignment of Sgt1 Against Other TPR Domain-containing Proteins** – **(A)** Alignment of Sgt1 homologues from yeasts using the BLOSUM62 scoring matrix against the sequences of the consensus three-repeat TPR structure (CTPR3, PDB ID 1NA0) and its triple-tyrosine mutant counterpart, CTPR3Y3 (PDB ID: 2WQH). This sequence alignment highlights the presence of a capping helix at the C-terminal end of the putative three-repeat TPR domain of Sgt1. In addition, it highlights two potential modes for self-association: an end-on C-terminally mediated dimerisation (■) and a loop-mediated stacking interaction brought about by intra-TPR loop tyrosines (▼). The alignment image was generated using ESPript (Gouet *et al.*, 1999). **(B)** An alignment of Sgt1 with the yeast cell cycle checkpoint protein Bub1 provides evidence for an electrostatic interaction between adjacent protomers via the inter-TPR loops (**Dashed Boxes**). Alternatively, a similar stacking interaction might also involve the hydrophobic residues found within these loops.

In order for the capping helix to be relevant in oligomerisation, it must be constitutively displaced from the last TPR repeat to expose residues critical for the oligomerisation interface. According to the sequence alignment of Sgt1 homologues from the yeasts *S. cerevisiae*, *K. lactis, C. glabrata* and *Ashbya gossypii*, with the sequence of the engineered consensus TPR structures CTPR3 and CTPR3Y3 (Figure 3.2), an equivalent hydrogen bond as in CTPR3 is unlikely to exist. In Sgt1 the equivalent residue-pair comprises a tryptophan and an aspartate, akin to the tyrosine and asparagine of CTPR3Y3, the structure of which shows displacement of the capping helix (Figure 3.2, ■) (Krachler *et al.*, 2010). Therefore, the TPR domain of Sgt1 may have the capacity to associate end-on with another protomer. In this mode of self-association, the 3B helices of each protomer would associate in parallel (as in CTPR3Y3) and therefore the C-termini would extend in the opposite directions on a single face of the associated TPRs. This would involve tight hydrophobic packing of 3B helices from adjacent protomers. This alignment also suggests that Sgt1 TPR protomers may form a stacking-type interaction via their loop regions. Sgt1 has tyrosines positioned in the intra-TPR loops that might support a mode of dimerisation similar to similar to Fis1 (Zhang *et al., 2*007) and CTPR3Y3 (Figure 3.2, A, ▼).

Sgt1 may also exhibit an electrostatic mode of self-association involving the inter-TPR loop regions, as evidenced by the alignment with the yeast protein Bub1 (Bolanos-Garcia *et al., 2*009) (Figure 3.2, B, dashed boxes). The inter-TPR loops denoted by this alignment also contain hydrophobic residues between TPR1B and TPR2A, and TPR3B and the capping helix, which could alternatively make this type of interface predominantly hydrophobic.

3.2 Design and Expression Trials of Protein Constructs for Biophysical Analysis and Crystallographic Studies

3.2.1 Design of Sgt1 Constructs

Using the information from the primary sequence and secondary structure analysis described above, and a local sequence alignment of Sgt1 homologues

(Figure 3.3 and Appendix A.1), a series of constructs were generated that could facilitate the delineation of the hydrodynamic properties of Sgt1, in particular the N-terminal TPR domain responsible for self-association. In addition, these constructs were used for crystallographic studies of Sgt1 and Sgt1:Skp1 complexes (see Chapter 4).

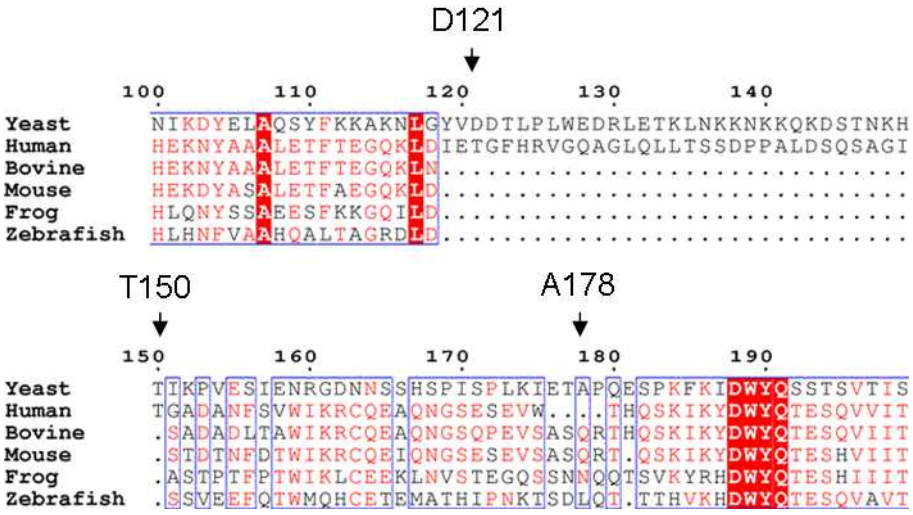


Figure 3.3: **Sequence Analysis of Sgt1 Homologues** – An alignment of Sgt1 homologues using the BLOSUM62 scoring matrix brought to light a sequence disparity in the C-terminal region of the TPR domain. This disparity was used to generate constructs for purification, and biophysical and structural characterisation. Sequence identity in the N-terminal region is low, in comparison to other regions of Sgt1 (data not shown). The alignment image was generated using ESPript (Gouet *et al.*, 1999).

Local sequence alignment of Sgt1 homologues using the BLOSUM62 scoring matrix underlined three separate targets for purification. The alignment produced a significant gap between the yeast sequence and other homologues (Figure 3.3). Despite low sequence identity between species, this gap was interpreted to indicate fundamental differences in the secondary structure of the N-termini of Sgt1, and accordingly aspartate 121 and threonine 150 were chosen as the C-terminal ends of two constructs, which respectively excluded and included this region of sequence disparity.

Based on a secondary structure prediction (Figure 3.1, A) and an assessment of the domain boundary with the CS domain (from the crystal structure of the Sgt1 CS domain), alanine 178 was chosen as a third cut-off. Sgt1 1-121, 1-150 and 1-

178 were constructs designed with crystallisation in mind (see Chapter 4), but would also provide information on the contribution of different sequences within the N-terminal TPR domain to the solution behaviour of Sgt1.

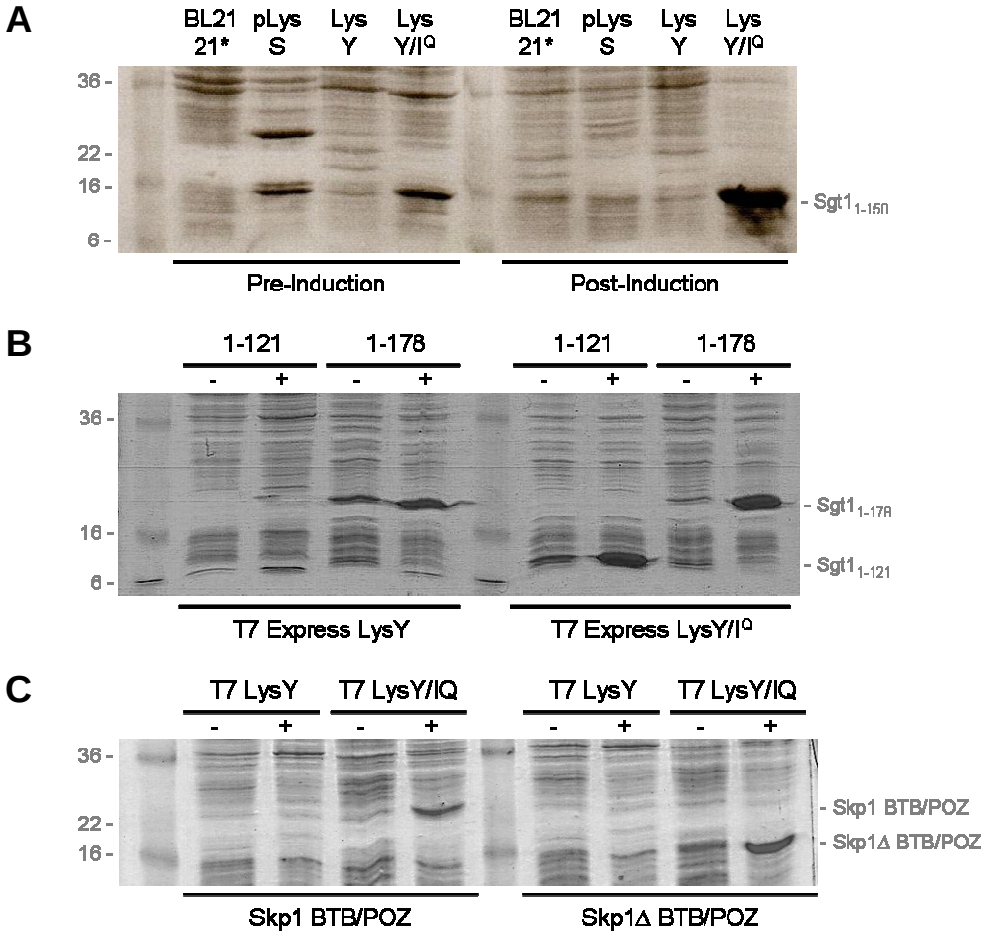


Figure 3.4: **SDS-PAGE Analysis of Expression Trial Cell Lysate Samples Pre- and Post-Induction –** All Sgt1 and Skp1 constructs for co-crystallisation were subjected to an expression trial to identify the most suitable *E. coli* derivative for expression. In each case, the NEB supplied cell strain T7 Express LysY/IQ provided the highest yield of recombinant protein expression. (A) Expression study of Sgt1 1-150, (B) Sgt1 1-121 and Sgt1 1-178, (C) Skp1 BTB/POZ and Skp1Δ BTB/POZ.

These constructs were sub-cloned into pRSETa using NdeI and BamHI restriction sites for tagless expression, using constructs previously generated in the group. For each construct, the T7 Express LysY/I^Q (NEB) strain (a derivative of BL21 cells with a chromosomally integrated lysozyme mutant for enhanced suppression of leaky expression) provided significant yields of protein (Figure 3.4, A and B), and it was consequently chosen as the host for large scale expression of the Sgt1 truncations.

3.2.2 Design of Skp1 BTB/POZ Domain Constructs

The BTB/POZ domain of Skp1 is a protein-protein interaction motif involved in the interaction with Sgt1. A structural alignment of all existing Skp1 structures deposited in the PDB identified the cut off for this minimal domain (Figure 3.5).

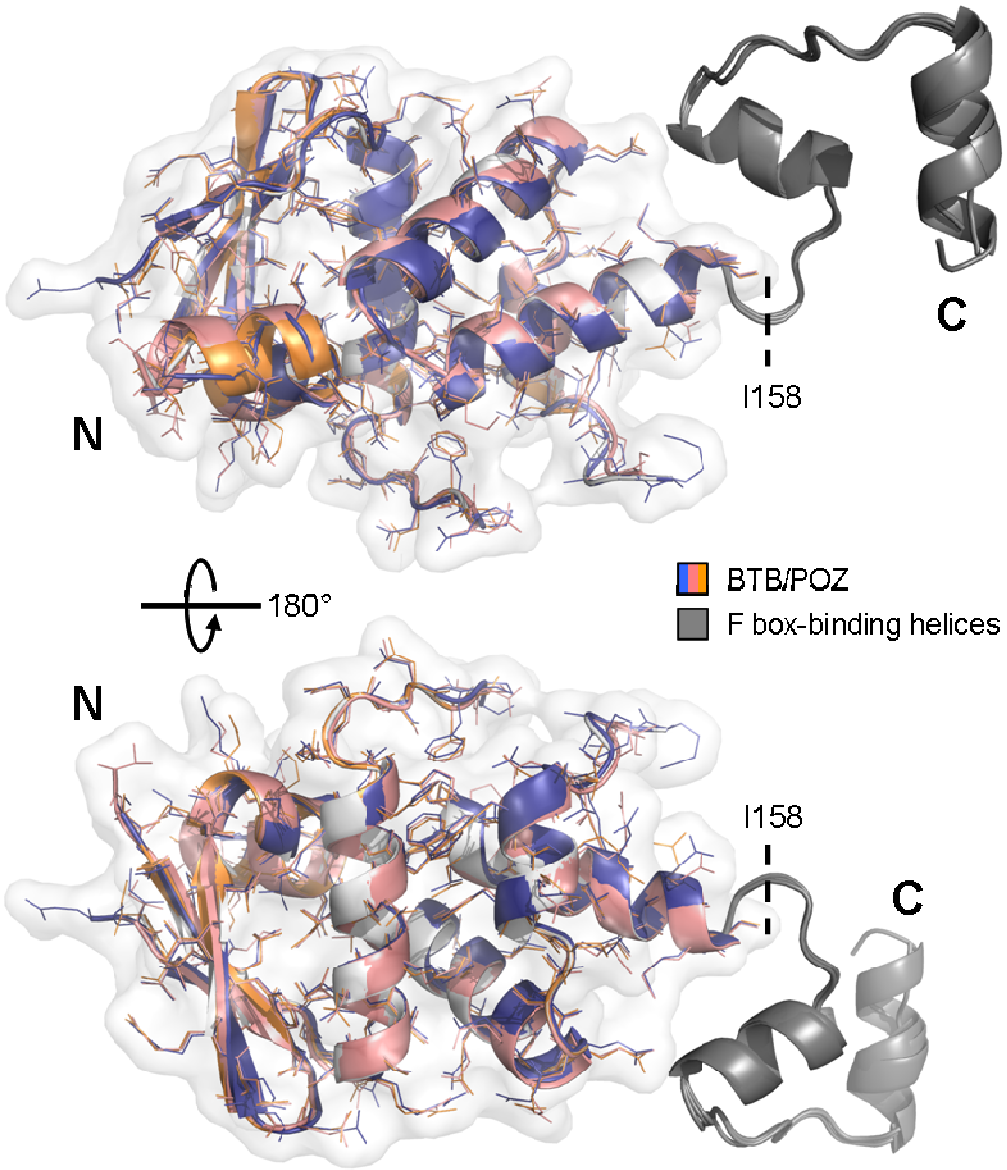


Figure 3.5: **Domain Boundaries of full length Skp1 from *Saccharomyces cerevisiae*** – A structural alignment of all yeast Skp1 structures (3MKS (**Dark Blue**), 1NEX (**Light Orange**) and 3V7D (**Salmon**)) deposited in the Protein Data Bank highlighted the boundary for a compact BTB/POZ domain at isoleucine 158. The BTB/POZ domain is shown in cartoon and line representation in colours corresponding to the respective PDB ID, with a surface envelope surrounding the domain to emphasize its compactness. Two loops are missing from each structure, approximately between residues 35-74 and 104-112. The F box-binding helices are shown in Grey.

The C-terminus of Skp1 comprises a conserved F box motif binding region consisting of three helices that cradle the F box component of the SCF ubiquitin ligase machinery (Figure 3.5, grey cartoon). The phenotype of the Skp1 kinetochore mutant, *skp1-4*, is suppressed by overexpression of exogenous Sgt1 (Bansal et al., 2004) and has a mutation in the BTB/POZ domain at leucine 146 (mutated to a serine). As this residue occurs in the core of the domain, it likely affects the fold and therefore destabilises the binding site for Sgt1 *in vivo*. Under the assumption that the interaction with Sgt1 therefore involved a region within the BTB/POZ domain, a construct lacking the C-terminal F box binding helices was generated, terminating in Ile 158.

All deposited structures of Skp1 in the context of SCF complexes (e.g. PDB ID 1FQV) do not contain information on the loop region between residues 35-74, which is predicted to be disordered (data not shown). The BTB/POZ construct was therefore generated from both wild type Skp1 and a previously cloned loop-less Skp1 by PCR (Skp1Δ), in order to help determine the role of this loop in the interaction with Sgt1. Both constructs were subsequently cloned into pRSETa (Invitrogen) via compatible NdeI and BamHI restriction sites for tag-free expression. Similar to the Sgt1 TPR constructs described previously, an optimization study on the expression of both genes highlighted significant expression in T7 Express LysY/I^{Q} (NEB) (Figure 3.4, C).

3.2.3 Purification of Sgt1 and Skp1 Constructs

As all designed constructs were to be purified without the use of an affinity tag, the protocol for purifying each was designed to capitalise on the unique properties of each protein. One crucial aspect of this was the need for reducing agent to prevent the formation of an artifactual disulphide bond between two Skp1 monomers, which were shown to form by comparing the band profile of purified Skp1 by SDS-PAGE under reducing and non-reducing conditions (data not shown). While the effect of this on Sgt1 binding was never tested, it seemed vital to maintain a monomeric form of Skp1 to bind a dimer of Sgt1. Furthermore, the presence of reducing agent seemed to have no effect on the ability of Sgt1 to

dimerise (see Section 3.3), meaning that active reduction of disulphide bonds would have no effect on the formation of a 2:1 stoichiometric complex. A summary of the result of these purification steps can be seen in Figure 3.6.

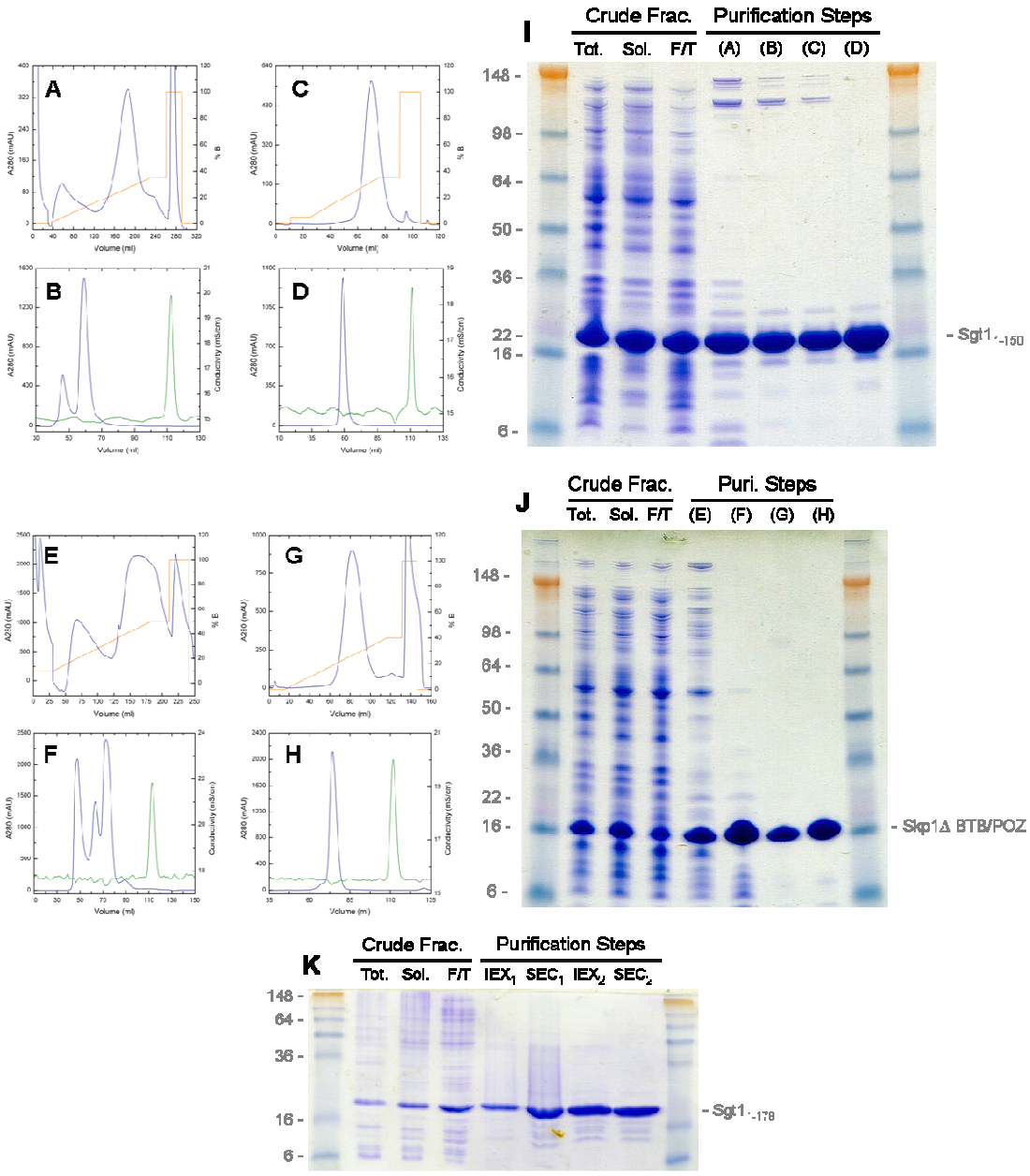


Figure 3.6: **Step-wise SDS-PAGE Analysis of Purification Methods for Biophysical Characterisation and Crystallisation Targets** – Purification of all constructs for crystallisation proceeded via four steps (two ion exchange and two size exclusion chromatography steps). Final purity was estimated at 95% for each gene product. **(A, E)** Chromatogram of first ion exchange step for Sgt1 (10 mL HiTrap SP FF) and Skp1 (10 mL HiTrap Q FF) constructs respectively. **(B, F)** First size exclusion chromatogram for Sgt1 and Skp1 (HiLoad 16/60 S75). **(C, G)** Second ion exchange step using a shallower gradient elution for enhanced separation of contaminating products (5 mL HiTrap SP/Q FF). **(D, H)** Final size exclusion chromatography step into 25 mM HEPES, 150 mM NaCl, 2 mM DTT, 1 mM EDTA, pH 7.0. Gels are shown for Sgt1 1-150 **(I)**, Skp1ΔBTB/POZ **(J)** and Sgt1 1-178 **(K)**, considered the most suitable targets for co-crystallisation.

Sgt1 and Skp1 constructs were purified based on their binding to cation and anion exchange chromatography columns, respectively (Figure 3.6). Constructs of both proteins bind strongly to their respective columns, though Skp1Δ BTB/POZ co-elutes with numerous bacterial proteins (Figure 3.6, E) that are only removed in subsequent steps. By virtue of their unique preference for a cation exchanger, Sgt1 constructs are well separated from contaminants after the first step (Figure 3.6, A).

Further purification produced samples with greater than 90% purity, with some proteolytic breakdown evident of the Sgt1 1-178 product. Not shown are the purification attempts for Sgt1 1-121: while the gene is overexpressed in the LysY/I^Q cell type, the product is entirely insoluble. It is possible that this truncation lacks a crucial component to enhance the solubility or maintain the fold of the TPR domain of Sgt1. Despite the fact that this construct would have likely been the most amenable to crystallisation with the BTB/POZ domain of Skp1 (by being smaller and presumably more compact), crystallisation attempts (see Section 4.1) and biophysical characterisation proceeded with Sgt1 1-150 and Sgt1 1-178.

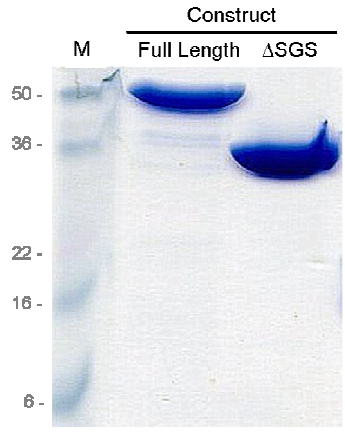


Figure 3.7: **SDS-PAGE Analysis of Purified Longer Constructs of Sgt1** – Protein constructs utilised in this study were purified by different chromatographic methods to 99% purity. The molecular weight ladder (**M**) shows that these proteins behave as expected in SDS-PAGE.

In addition to these TPR-centred constructs, a truncation of Sgt1 lacking the C-terminal SGS domain and full length Sgt1 were purified successfully, and used exclusively for biophysical characterisation. A summary SDS-PAGE analysis of the purified products can be seen in Figure 3.7. Untagged full length Skp1 was

expressed and purified to a similar level of purity and shown to be clipped at the N-terminal methionine by mass spectrometry (data not shown).

3.3 Biophysical Characterisation of Sgt1

As previously discussed, it has been reported that barley Sgt1 contains two cysteines within the TPR region that form an intramolecular disulphide bond and regulate oligomerisation (Nyarko *et al., 2*007). The TPR domain of Sgt1 contains a single cysteine, which, in the absence of a reducing agent, could lead to artifactual self-association. To probe for the role of this cysteine in TPR mediated dimerisation, SV-AUC analyses of Sgt1 were carried out in both the presence and absence of 2 mM DTT. Secondly, to probe the nature of the Sgt1 dimer equilibrium in solution, samples of full length Sgt1 (in the presence of reducing agent) exposed to a prolonged 24 hour equilibration at 4°C, after dilution from stock to working concentrations, were compared to samples that had not been subject to equilibration (and were therefore diluted immediately prior to analysis). This experiment would therefore account for a slow equilibrium with metastable oligomers that would be present only before re-equilibration at a lower sample concentration. Sedimentation was monitored at 42000 RPM at 20°C using both absorbance and interference optics over a period of approximately 72 hours.

3.3.1 Sedimentation Velocity Analytical Ultracentrifugation of Full Length Sgt1

Data analysis using a continuous c(s) distribution model gave rise to a two peak distribution of species (Figure 3.8). The main species has a sedimentation coefficient ranging from 3.8 to 4.3 S depending on concentration. This information can be extrapolated to a weight-average molecular weight estimate of approximately 93400 Da at an equilibrated sample concentration of 22.3 μM (1.0 mg/ml) and 79900 Da at the lowest concentration of 4.5 μM (0.2 mg/ml), and a similar molecular weight estimate for the non-equilibrated sample.

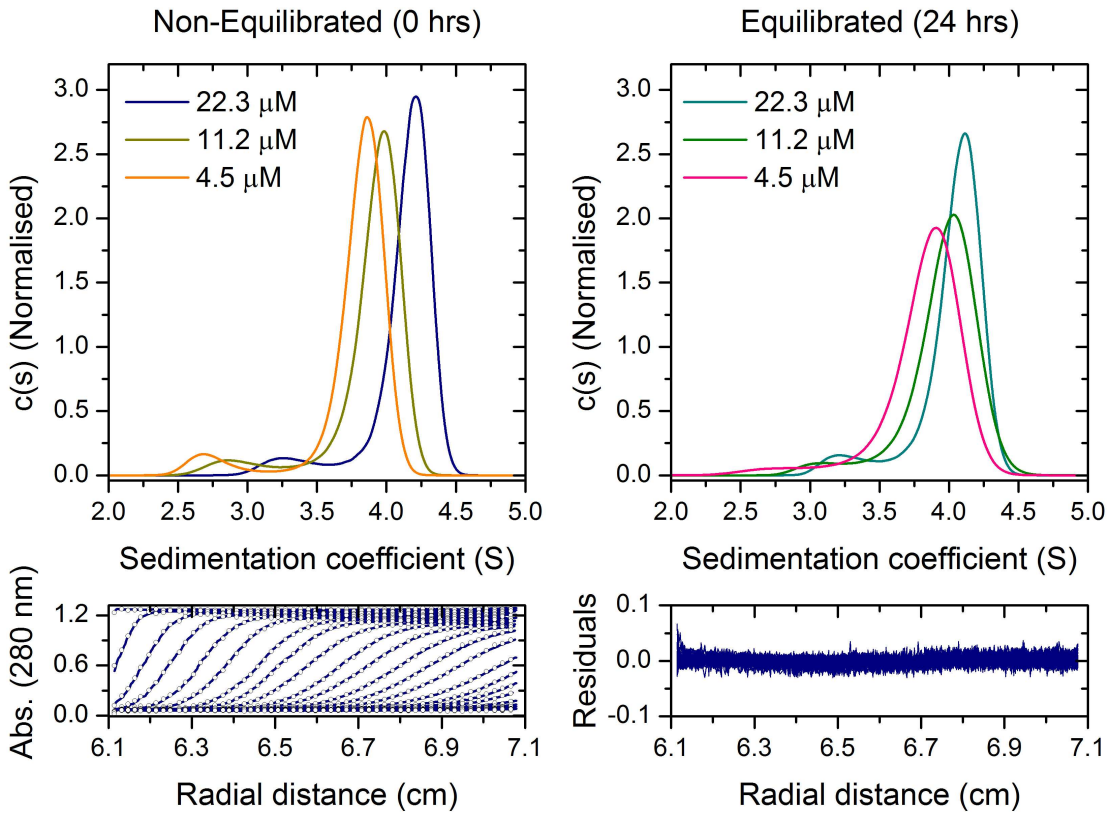
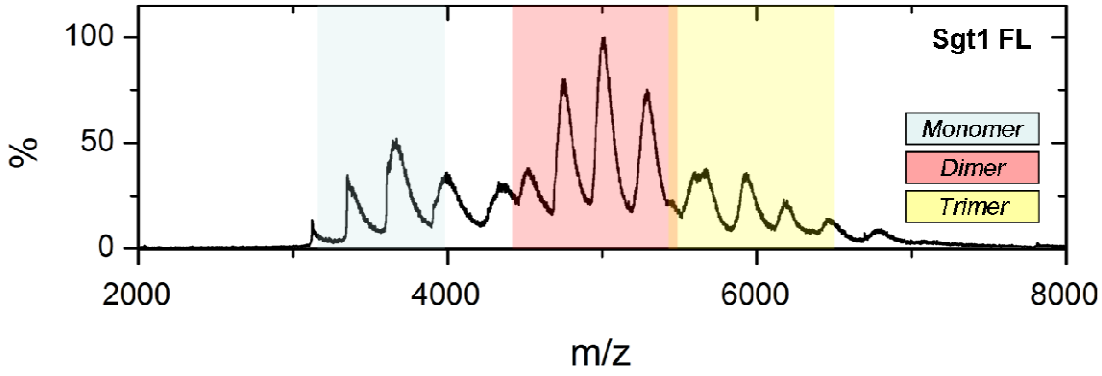


Figure 3.8: **Sedimentation Velocity Analytical Ultracentrifugation of Full Length Sgt1 in the Presence of Reducing Agent** – Samples were prepared at concentrations ranging from 4.5 – 22.3 µM (equivalent to 0.2 to 1.0 mg/ml). Samples were buffer exchanged by size exclusion into 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM DTT and 1 mM EDTA prior to analysis. Sedimentation was carried out at 42000 RPM at 20°C for approximately 72 hours. This figure summarises these results, with a representative fit derived from the pre-equilibrated sample at 22.3 µM protein concentration (**Bottom**). The initial profile represents the distribution plot of peaks corresponding to species identified by the best-fit analysis of absorbance data collected at a wavelength of 280 nm, with the **Top Left** plot being from the non-equilibrated and the **Top Right** from the equilibrated samples.

A dimer of Sgt1 has an expected molecular weight of approximately 89457 Da. Therefore, the molecular weight estimate based on a weight-average frictional ratio provides evidence for the predominance of a dimer species with and without equilibration.

This is in line with previous sedimentation equilibrium (SE) experiments highlighting strong Sgt1 dimerisation in solution (Bansal *et al.., 2*008). The distribution also reveals a second minor species, not visible in the SDS-PAGE analysis (Figure 3.7). This species is attributed a molecular weight corresponding to a monomer of 45000 Da, with a sedimentation coefficient between 2.5 and 3.1 S. The distribution is unaffected by the absence of reducing

agent (data not shown), which implies that self-association is not regulated by disulphide bond formation. The weight-average frictional ratio (1.7 for equilibrated and 1.8 for non-equilibrated samples at the highest concentration) suggests that the species in solution are elongated. The lack of change in s-value of the main species between the two sets of data argues that there are no metastable populations that might disappear in the analysis after equilibration.



Species	Stoichiometry	Expected Mass (Da)	Observed Mass (Da)	m/z Range
Sgt1 FL	1	45886	46867 ± 1.66	3348.64 – 3906.58
	2	91772	93734 ± 3.32	4464.52 – 5514.76
	3	137658	140601 ± 4.98	5408.73 – 6391.95

Figure 3.9: **Native Electrospray Ionisation Mass Spectrometry Analysis of Full Length Sgt1** – Mass spectrometry analysis of full length Sgt1 shows that monomer, dimer and trimer solution species can be detected in the gas phase. The predominant species is a dimer, mirroring the SV-AUC solution analysis of Sgt1. Spectra collection and data analysis was carried out by Richard Kerr.

There are two possible rationales for a two-peak distribution, if what is observed is related to a monomer:dimer equilibrium. First, it is possible that the published K_D for a dimer of Sgt1 (Bansal *et al*, 2008) is incorrect and that the two-peak distribution represents a slow equilibrium of monomers and dimers. This would mean that Sgt1 actually has a K_D in the experimental concentration range and that slow kinetics allows the resolution of species on the time scale of the experiment. However, since the main species shifts with concentration in both sets of data, there is some fundamental concentration-dependent change in the equilibrium that cannot be explained by slow kinetics. Second, the distribution shows a 'dead' monomer, represented by the minor peak in the distribution, which doesn't exchange with the main dimer species. In this case, one might expect an unchanging peak centre but decreasing peak height as protein concentration is decreased. While the signal for this species is likely too low to

conclude reliably on this possibility (and therefore, also on the first rationale), the parallel decrease in abundance with concentration means that even through interpretation this explanation is not viable.

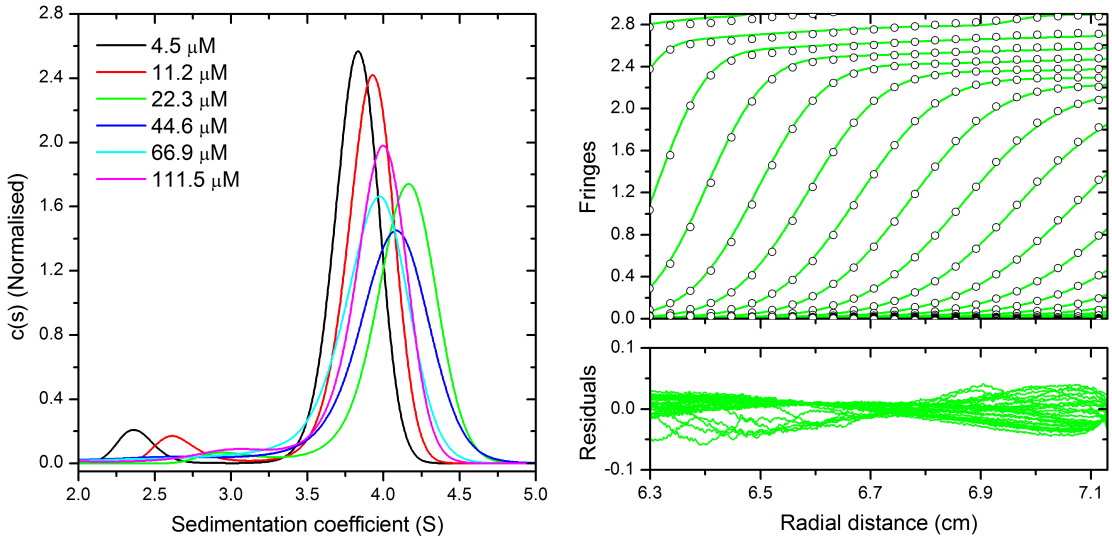
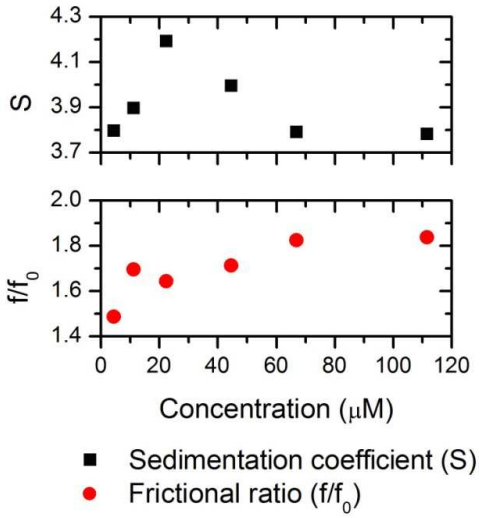


Figure 3.10: **SV-AUC of Full Length Sgt1 at Concentrations Greater Than 22.3 μM** – Higher concentration samples were prepared in an identical buffer to the previous samples (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM DTT and 1 mM EDTA). Sedimentation was monitored as before by both absorbance and interference optical systems. The example of raw absorbance data (collected at 280 nm) and residuals shown to the right-hand side is of the 22.3 μM sample (**Right**). The distribution plot (**Left**) that the predominant population is a dimer between 3.5 and 4.5 S. The unusual shift in sedimentation coefficient at concentrations greater than 22.3 μM is likely caused by non-ideal behaviour at high sample concentrations.

Therefore, assuming that the published low nanomolar K_D is correct and that any interpretation of the smaller species is unreliable, the only reasonable explanation for these trends is that the distribution reflects an equilibrium of oligomers beyond a dimer and that, at the highest concentration, the observations are of the top end of a secondary self-association equilibrium (from dimer to trimer or greater). If this is the case, then the higher order equilibrium exhibits fast kinetics (a high k_{off}) based on the concentration-dependent shift in the s-value of the main species. This explanation maintains a 20 nM K_D for the dimer, but a different K_D for any higher molecular-weight oligomers, and increases in concentration should therefore yield an equivalent increase in the estimated molecular weight. Evidence for this hypothesis is provided by native ESI mass spectrometry experiments, which show a small gas phase-detectable

population of full length Sgt1 trimers that co-exist with monomers and dimers in solution (Figure 3.9).



Conc. (mg/ml)	Conc. (μm)	s-value of Main Species	Best fit f/f_0
5.0	111.5	3.8	1.5
4.0	66.9	3.8	1.7
2.0	44.6	3.6	1.6
1.0	22.3	4.2	1.7
0.5	11.2	3.9	1.8
0.2	4.5	3.8	1.8

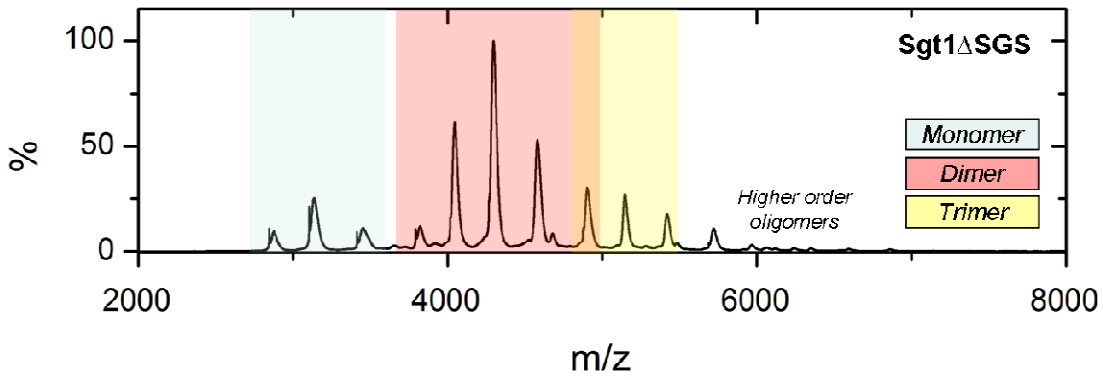
Figure 3.11: **Scatter Plot Relating Sedimentation Coefficient and Frictional Ratio to Concentration of Full Length Sgt1** – Scatter plot showing the distribution of sedimentation coefficient (S) (**Black Squares**) and frictional ratio (f/f_0) (**Red Circles**) with changes in concentration. The plot reinforces the idea of an erratic shift in S of the main species, as caused by non-ideal behaviour.

Further SV-AUC analyses were performed of full length Sgt1 to test the hypothesis of concentration-dependent higher-order oligomerisation (Figure 3.10). The main peak undergoes an evident shift in the sedimentation coefficient with concentration, though there is no consistent trend in the data: from 4.5 – 22.3 μM, the shift reflects an increase in the s-value from 3.8 to 4.2, though the s-value decreases at higher concentrations, down again to 3.8 at 111.5 μM (Figure 3.11). It is likely that this is caused by non-ideal behaviour of particles in solution. While the main species is still related to a dimer, the s-values at these higher concentrations are consequently unreliable indicators of the hydrodynamic properties of full length Sgt1.

3.3.2 Sedimentation Velocity Analytical Ultracentrifugation of Sgt1ΔSGS

It has been shown by *in vitro* phosphorylation studies and characterisation of phosphomimetic mutants that the C-terminus of Sgt1, the SGS domain, affects oligomerisation as a consequence of phosphorylation by casein kinase 2 (Bansal *et al.., 2*010). ESI-MS studies under native conditions suggest that the domains or sequences C-terminal to the TPR affect the level of self-association (Figure 3.12),

but that the SGS domain by itself does not hinder oligomerisation. Further SV-AUC studies were carried out on a construct lacking the C-terminal SGS domain (Sgt1ΔSGS) to understand the contributions of this domain to oligomerisation in solution.



Species	Stoichiometry	Expected Mass (Da)	Observed Mass (Da)	m/z Range
Sgt1ΔSGS	1	33207	34119.05 ± 21.95	*Data not available*
	2	66414	68242.34 ± 62.26	
	3	99621	102382.45 ± 50.76	

Figure 3.12: **Native Electrospray Ionisation Mass Spectrometry Analysis of Sgt1ΔSGS** – Gas phase analysis of Sgt1ΔSGS shows a predominance of dimeric species of Sgt1 in solution, which can be detected in the gas phase. Truncation of the C-terminal SGS domain results in an apparent decrease in the relative population of monomers. Spectra collection and data analysis was carried out by Richard Kerr.

The peak distribution shows a small decrease in s-value of the main species from 3.5 to 3.3 S at the lowest concentration of 3.1 μM (Figure 3.11). The average frictional ratio across all concentrations is 1.6 and there is little or no change between concentrations (Figure 3.13). From its primary sequence, the monomer has an expected molecular weight of 31978 Da. Extrapolating this information to a weight-average molecular weight estimate for the main species suggests that it is a dimer, with a molecular weight of 63957 Da at the highest concentration tested. In this respect, the ΔSGS construct behaves similarly to the full length protein in solution, in that the main species is a dimer.

In contrast, however, there are no apparent lower s-value species of Sgt1ΔSGS. Instead the best-fit analysis reveals a species with a larger sedimentation coefficient between 5.7 and 6.4 S (Figure 3.13, sub-peak).

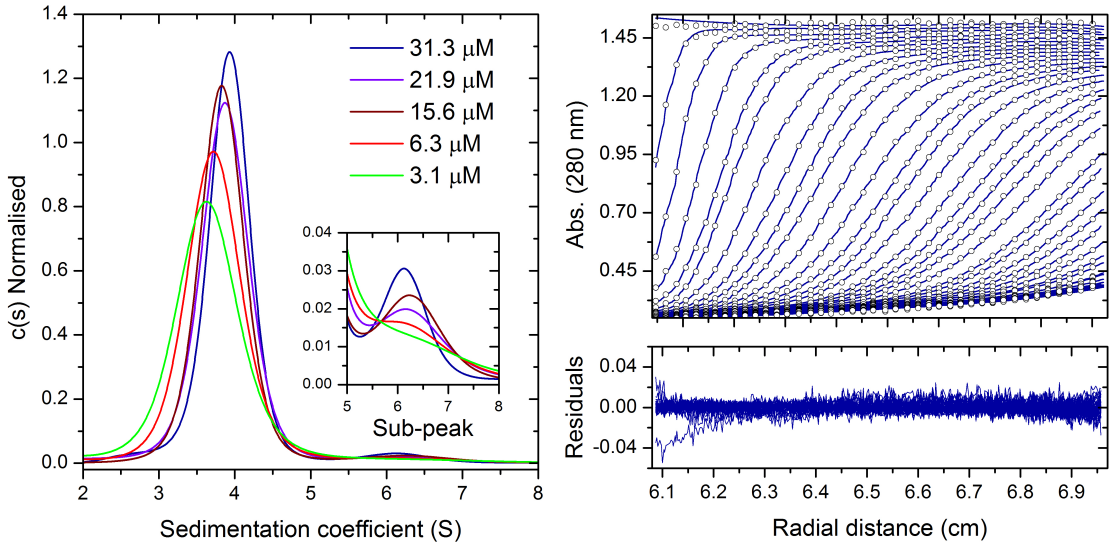
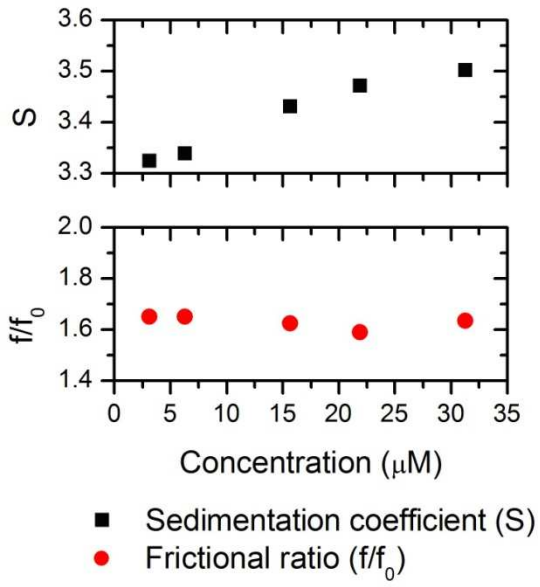


Figure 3.13: **Sedimentation Velocity Analytical Ultracentrifugation of Sgt1ΔSGS at Different Concentrations**– The ΔSGS protein was purified by chromatographic means and buffer exchanged into 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM DTT and 1 mM EDTA prior to analysis. Protein concentrations were confirmed by absorbance at 280 nm following pre-equilibration of sample for 24 hours. The fit and residuals data corresponds to the 31.3 μM sample and is representative of the general error-level of the absorbance datasets obtained at a wavelength of 280 nm (**Right**). The distribution plot shows a two peak distribution profile, with the predominant species having a lower s-value compared to the higher species identified by SEDFIT (**Left**). This suggests that deleting the SGS domain influences dimerisation and allows concentration-dependent higher order oligomerisation; a small population of tetramer appears to be present, with a non-shifting sedimentation coefficient (any minor shifts can likely be attributed to noise). The largest population represents a dimer of Sgt1ΔSGS with a shift in sedimentation coefficient. We attribute this change to be a concentration effect.

The molecular weight estimate for this is consistently about twice the molecular weight of the main dimer across three of the five concentrations tested and is possibly a tetrameric species. A trimeric species of Sgt1ΔSGS could also be detected by mass spectrometry (Figure 3.12); however, with both main species accounted for, there was no evidence for a trimer species in solution.

While the s-value of the main peak decreases with a decrease in concentration, that of the sub-peak shifts to a higher s-value (Figure 3.14, *table*). This latter behaviour is probably the result of systematic noise as the abundance of the second species is significantly lower than the dimer and therefore the absorbance data becomes unreliable for this species. The change in s-value of the main peak with concentration and the insignificant change in frictional ratio suggests that the shift in the weight-average molecular weight is caused by the s-value alone and not the frictional ratio.



Conc. (mg/ ml)	Conc. (μm)	S of Sp. 1	S of Sp. 2	f/f_0
1.0	31.3	3.5	5.7	1.6
0.7	21.9	3.5	6.2	1.6
0.5	15.6	3.4	6.4	1.6
0.2	6.3	3.3	n/a	1.7
0.1	3.1	3.3	n/a	1.7

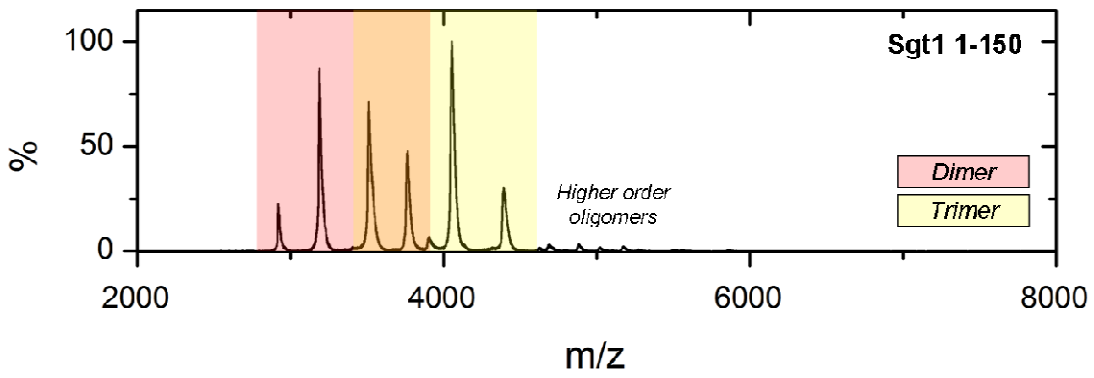
Figure 3.14: **Scatter Plot Relating Sedimentation Coefficient and Frictional Ratio to Concentration of Sgt1ΔSGS** – The concentration:sedimentation coefficient correlation of 'Species 1' (the main species) is represented by **Black Squares**, while the corresponding frictional ratio:concentration relationship is depicted by **Red Circles**. Decreasing concentration of the ΔSGS leads to a decreasing trend in sedimentation coefficient of the main species, while the weight-average frictional ratio determined by the fit effectively remains constant. By contrast, the second species increases in sedimentation coefficient, before becoming undetectable through the data fitting at a concentration of 6.3 μM. ***Sp*** = Species.

3.4 Biophysical Characterisation of the N-terminal TPR Domain of Sgt1

Section 3.1 describes features in the primary sequence of the Sgt1 N-terminus characteristic of oligomerising TPRs. ESI mass spectrometry experiments have illustrated that Sgt1 TPR domain constructs can form higher order oligomers in 150 mM ionic strength buffer that could be detected in the gas phase (Figure 3.15). To corroborate the mass spectrometry experiments, solution studies of Sgt1 1-150 and Sgt1 1-178 were carried out in 150 mM ionic strength buffer.

3.4.1 Sedimentation Velocity Analytical Ultracentrifugation of Sgt1 1-150

Sedimentation velocity experiments of Sgt1 1-150 were carried out under similar conditions as described above, with the exception of being sedimented at 60000 RPM due to the lower molecular weight of the analyte.



Species	Stoichiometry	Expected Mass (Da)	Observed Mass (Da)	m/z Range
Sgt1 1-150	2	34924	34663.66 ± 6.46	2889.64 – 3852.52
	3	96330	51995.49 ± 9.69	3467.37 – 4727.86

Figure 3.15: **Native Electrospray Ionisation Mass Spectrometry Analysis of Sgt1 1-150** – Native ESI-MS analysis of Sgt1 1-150 (a variant truncation of the TPR domain) does not detect a monomeric species in the gas phase. Instead, the predominant species (or equilibria) that are transferred from solution appear to be dimers and trimers of Sgt1 1-150. Spectra collection and data analysis was carried out by Richard Kerr.

A range of concentrations were prepared to probe for concentration-dependent oligomerisation. At the highest concentration tested, the best-fit analysis using a continuous c(s) distribution model reveals two peaks with sedimentation coefficients of approximately 2.6 and 1.9 S, and a frictional ratio of 1.8 (Figure 3.16).

Decreasing concentrations diminish the minor population, with only a slight shift in the s-value of the main species to 2.5, 2.6 and 2.6 S at 57.7, 40.4 and 28.9 μM concentrations respectively (Figure 3.16 and Figure 3.17). The slight fluctuation in the s-value is likely due to systematic noise and is therefore insignificant. The expected monomer molecular weight of Sgt1 1-150 is approximately 17331 Da, while the extrapolated weight-average molecular weight for the main peaks in all samples has a mean of 51992 Da. This shows that Sgt1 1-150 can form trimers in solution, which corroborates the ESI-MS analysis, and infers that the main population in samples of Sgt1 1-150 is a dimer:trimer equilibrium.

There is no significant change in the s-value at the lowest concentration of 17.3 μM (2.5 S), though there is a significant change in the relative peak width (Figure 3.16). While the fitted frictional ratio should be disregarded due to the low

signal:noise ratio at this concentration of protein (equivalent to 0.3 mg/ml), this change in peak shape is indicative of a dissociation event occurring between 28.9 and 17.3 μM. If this is the case, then these analyses support the previous hypothesis that there might be differences in the affinity of association of monomers into dimers, and dimers into higher order oligomers.

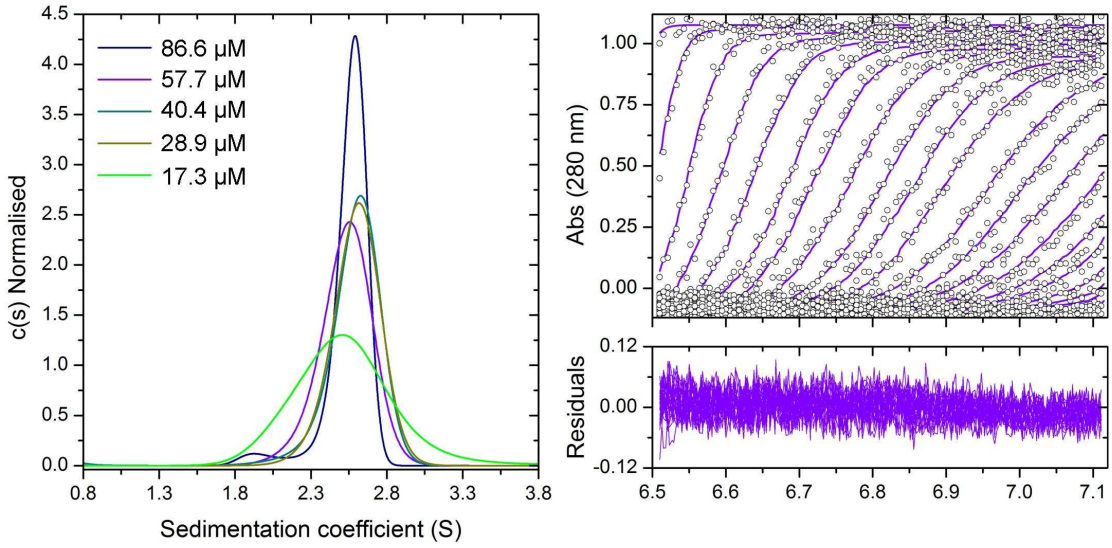
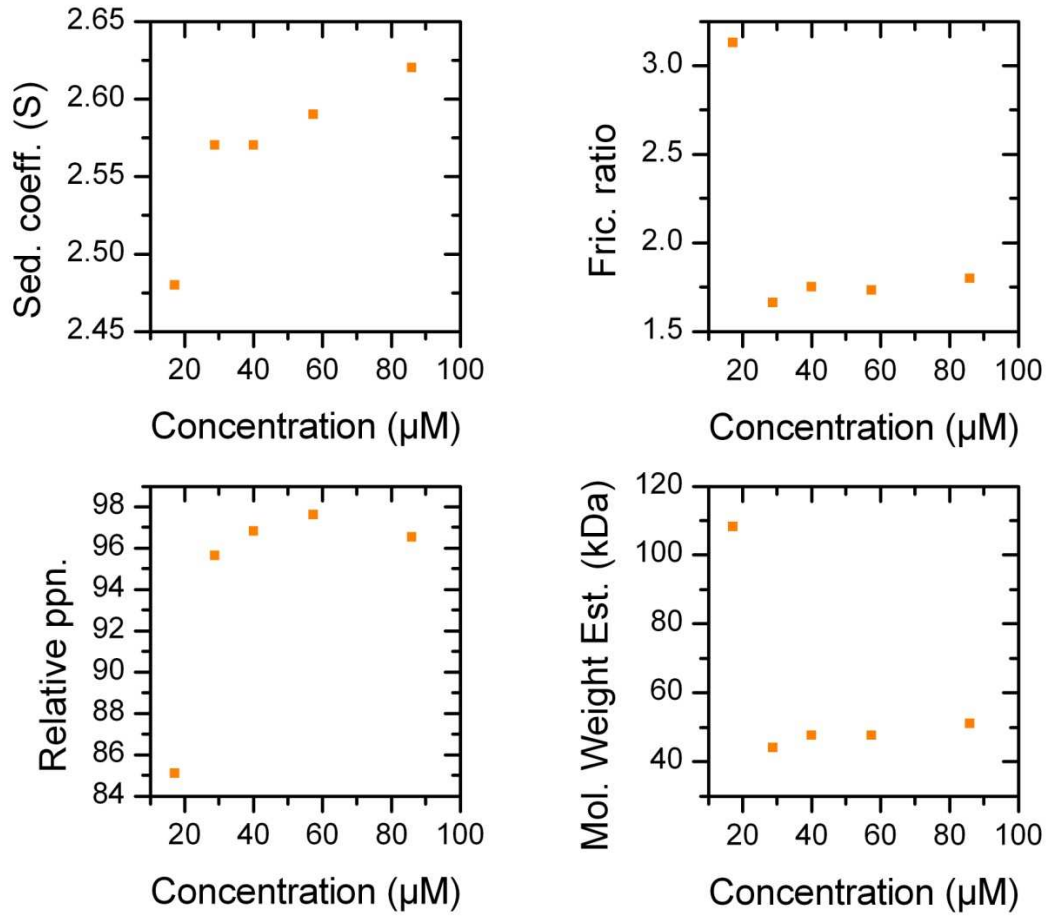


Figure 3.16: **SV-AUC of Sgt1 1-150 at Different Concentrations** – Samples were prepared in 20 mM HEPES pH 7.0, 150 mM NaCl, 2 mM DTT and 1 mM EDTA at the appropriate concentration, as confirmed by absorbance at 280 nm. The raw data and residuals information (**Right**) were derived from the 86.6 μM sample, and are representative of absorbance scans collected at a wavelength of 280 nm. The distribution plot (**Left**) depicts a large single peak corresponding to a trimer. At the highest concentration, a secondary peak is also present. Decreasing the concentration of this species to 17.3 μM has a significant effect on the frictional ratio and therefore the weight-average molecular weight estimate, suggesting that the trimer is affected by concentration. All species have a sedimentation coefficient of approximately 2.5 S.

3.4.2 Sedimentation Velocity Analytical Ultracentrifugation of Sgt1 1-178

Sgt1 1-178 comprises the core TPR domain and additional unstructured elements C-terminal to the capping helix. In order to determine whether the additional length of sequence attributes a different solution behaviour to this construct, purified samples of 1-178 were subjected to SV-AUC to delineate its hydrodynamic properties. Sample analysis was carried out as for the 1-150 construct. The main species, at concentrations from 24.6 – 73.7 μM (equivalent to 0.5 – 1.5 mg/ml), has a sedimentation coefficient of approximately 2.9 S (Figure 3.18), with a weight-average frictional ratio between 1.4 – 1.5, implying that the 1-178 is less elongated than previously characterised populations of

Sgt1 constructs. A dimer of Sgt1 1-178 should have a theoretical molecular weight of 40692 Da. The main species, with a mean weight-average molecular weight of 46406 Da, is therefore slightly larger than anticipated for a dimeric species. The smaller weight-average frictional ratio compared to Sgt1 1-150, however, supports the idea that 1-178 has a smaller hydrodynamic radius and is therefore dimeric, suggesting that the extra sequences (C-terminal to 150) sterically hinder higher order oligomerisation.



Conc. (μM)	Sed. coefficient (S)	Frictional ratio	Relative population (%)	Molecular Weight Estimate (kDa)
85.9	2.6	1.8	96.5	51.0
57.3	2.6	1.7	97.6	47.6
40.1	2.6	1.8	96.8	47.6
28.7	2.6	1.7	95.7	44.0
17.2	2.5	3.1	85.1	108.3

Figure 3.17: **Scatter Plot Relating the Fitted Data Analysis Parameters to Concentration of Sgt1 1-150** – This correlation makes it clear that there is no significant change in s-value or frictional ratio with concentration. The fitted frictional ratio parameter of the lowest concentration should be disregarded due to a poor signal:noise ratio, but the peak is nonetheless relevant in connection to the hydrodynamic behaviour of Sgt1 1-150 at different concentrations. Sedimentation coefficients and frictional ratios listed in the table are rounded off to one decimal place.

A second minor species with a molecular weight estimate between monomer and dimer is detected in the continuous c(s) distribution (Figure 3.18). The distinction of this species from the main peak suggests that there are two non-interacting species that can be separated on the time-scale of the experiment. However, at a concentration of 9.8 μM, only a single predominant species is detected with a lower sedimentation coefficient of 2.7 S, reflecting the collapse of two species into a single interchanging population. Since the molecular weight estimate, based on a weight-average frictional ratio of 1.3, approximates a dimer (35765 Da), this single species likely represents a monomer-dimer equilibrium.

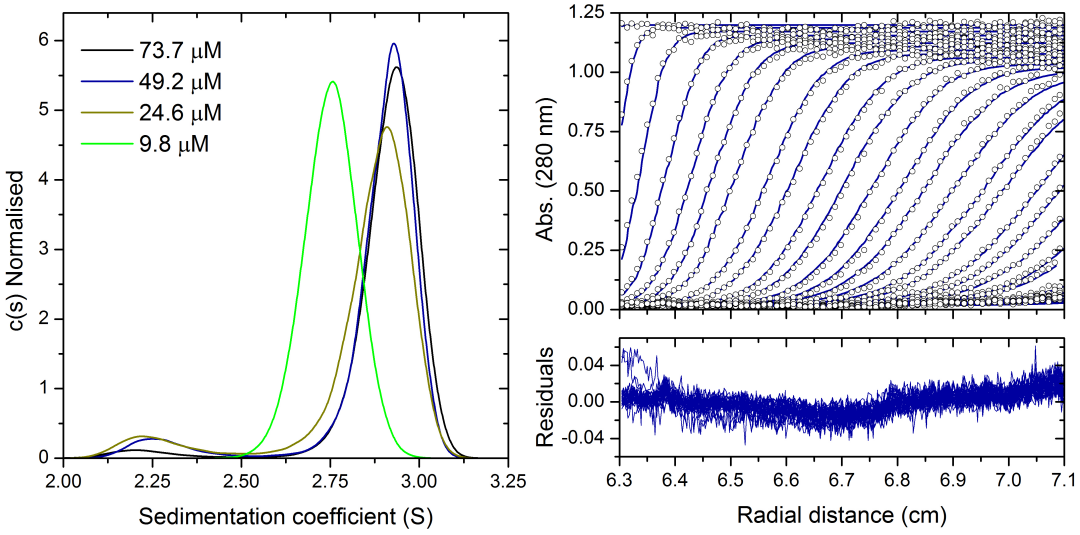


Figure 3.18: **SV-AUC Data Analysis of Sgt1 1-178 at Different Concentrations –** Sgt1 1-178 was purified as described in Section 3.2 and analysed in buffer containing 25 mM HEPES pH 7.0, 150 mM NaCl, 2 mM DTT and 1 mM EDTA. Protein concentrations were confirmed by absorbance at 280 nm following overnight sample pre-equilibration. The fit and raw data (**Right**) are taken from the data for 73.7 μM sample concentration and are representative of all absorbance data collected at a wavelength of 280 nm, showing a good fit of the raw data. The distribution profile (**Left**) shows a two peak distribution at concentrations greater between 24.6-73.7 μM (**Green**), with the equilibrium collapsing into a single peak distribution at 9.8 μM.

In order to optimise the molecular weight estimate for the populations present at concentrations greater than 9.8 μM, a bimodal frictional ratio model was applied to the data to improve the fitting parameters (Figure 3.19). This model attempts to fit a separate frictional ratio to a set range of s-values in order to improve the molecular weight estimate for species within the specified range. As such, it relies on the accuracy of an initial continuous c(s) distribution to specify a range of s-values specific to a species with distinct hydrodynamic properties, and not

necessarily part of the predominant equilibrium in solution. If the boundaries of two species are closely related, then this model cannot be used. For this reason, the bimodal frictional ratio model could not be applied to previous analyses of longer Sgt1 constructs with a two-peak distribution. However, if successful, then the fitting can highlight hydrodynamic differences in two related populations that might otherwise be subject to the bias of a weight-average interpretation. This was considered important in light of the observations of a higher order equilibrium beyond a dimer, since this might occur via a secondary self-association site that would result in two pathways to higher order oligomerisation, and two different dimer conformations.

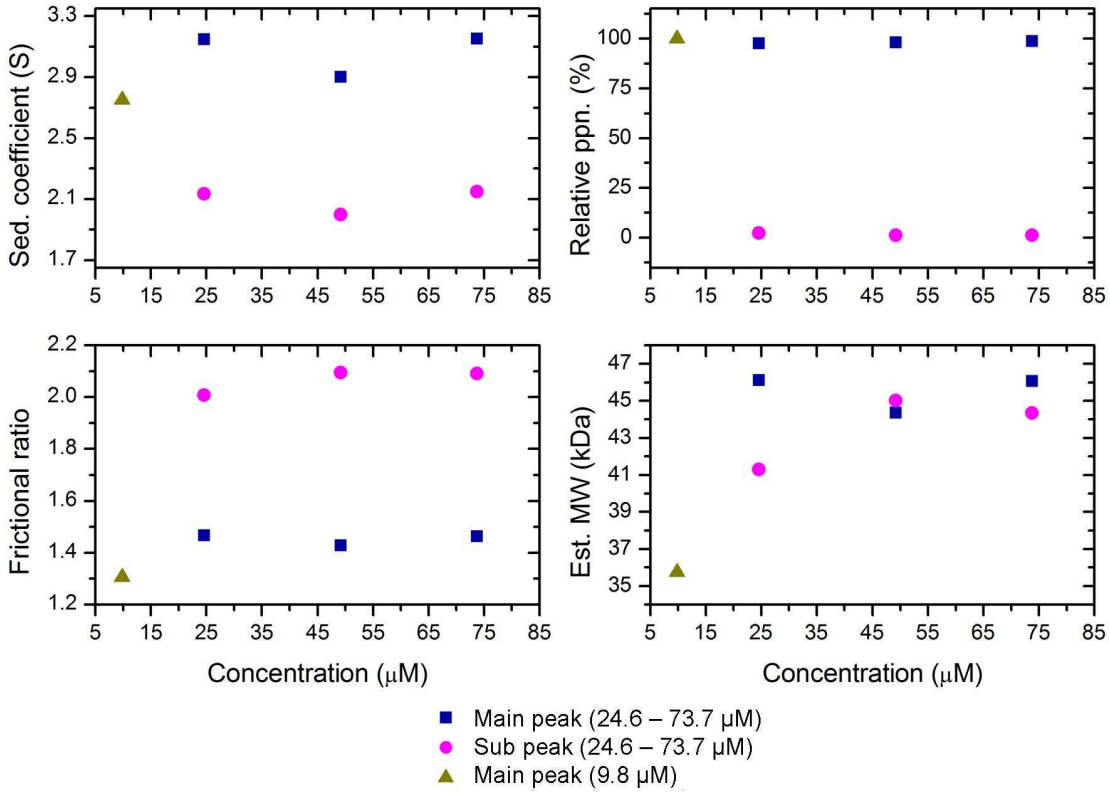


Figure 3.19: **Summary of Fitted Parameters of a Bimodal Frictional Ratio Model Applied to Sgt1 1-178 SV-AUC Absorbance Data** – Using a bimodal frictional ratio model highlights the possibility of two dimer:trimer equilibria of Sgt1 1-178 undergoing slow exchange in solution. The factors described show the change in sedimentation coefficient (**Top Left**), frictional ratio (**Bottom Left**), relative population (**Top Right**) and estimated molecular weight (**Bottom Right**), for both the major and minor species at concentrations at which both are present (24.6-73.7 μM). At the lowest concentration, only a single distribution is shown by the bimodal frictional ratio model, and approximates a dimer based on the weight-average frictional ratio and molecular weight estimate (**Red Triangle**). The two peaks reflect species with significantly different frictional ratios, with the more elongated of the two (**Magenta Circle**) comprising a relatively small proportion of the total population detected by the data fitting process.

The fitting of a bimodal frictional ratio model highlights a significant discrepancy in the hydrodynamic radius for both species at concentrations greater than 9.8 μM (Figure 3.19). With an average best-fit frictional ratio of approximately 1.5, the predominant species is expected to have a slightly elongated shape and a weight estimate slightly larger than a dimer (45516 Da). This compares favourably to the continuous c(s) analysis. In contrast, the second species has an average best-fit frictional ratio suggestive of a significantly extended macromolecule (2.06) and a molecular weight estimate of 43550 Da. The s-values for both species are consistent between both sets of analyses. This fitting model therefore suggests that both peaks represent dimeric populations, with the slightly skewed molecular weight estimate inferring that each might represent a dimer:trimer equilibrium.

It is clear that sequences between residues 150 and 178 of Sgt1 play a role in modulating self-association. Deletion of residues C-terminal to threonine 150 in Sgt1 permits further association, which can be detected by SV-AUC. At each of the highest concentrations tested for full length, ΔSGS and 1-178, an increase in the frictional ratio parallels an increase in molecular weight (Figure 3.20). At the same time, Sgt1 1-150 shows a significantly higher frictional ratio compared to the other constructs.

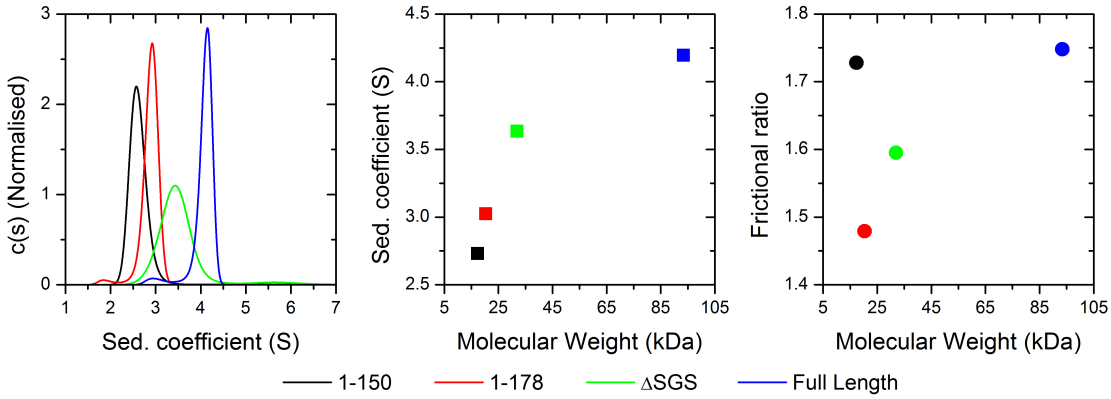


Figure 3.20: **Comparison of Sgt1 Constructs Analysed by SV-AUC** – As expected, an increasing trend in the sedimentation coefficient occurs with an increase in the monomeric molecular weight of the construct tested. Constructs encompassing residues 1-178 to 1-395 (full length) show a similar increasing trend in the frictional ratio, due to the linear addition of structural features that do not affect dimerisation. However, truncating Sgt1 to residues N-terminal to 178 (i.e. Sgt1 1-150) allows further oligomerisation beyond a dimeric species, which is accompanied by an increase in the weight-average frictional ratio attributed to the single distribution (**Far Right**).

3.5 Analysis of Sgt1 by Size Exclusion Chromatography-coupled Multi-angle Light Scattering

An analysis of the expected sedimentation coefficient for a trimer of Sgt1 1-136 using HYDROPRO (Ortenga *et al*., 2011) (applying the structure of Sgt1 discussed in Chapter 4) revealed that it would have an s-value of approximately 3.5 S, which differs significantly from the experimentally derived s-value for all TPR constructs of Sgt1 (Table 3.1).

Sgt1 Model	Bead radius (Å)	Sedimentation coefficient (S)	Rotational Diffusion Coefficient (s^{-1})	Translational Diffusion Coefficient (cm^2/s)
Monomer	2	1.8	$2.2 \cdot 10^{7}$	$1.1 \cdot 10^{-6}$
Dimer	2	2.8	$1.1 \cdot 10^{7}$	$8.6 \cdot 10^{-6}$
Trimer	2	3.5	$6.5 \cdot 10^{6}$	$7.3 \cdot 10^{-7}$

Table 3.1: **HYDROPRO Analysis of Sgt1 Oligomers –** HYDROPRO (Ortenga *et al*., 2011) calculates the theoretical hydrodynamic properties of an atomic model (i.e. a PDB file), by modelling the input structure as beads of different sizes, depending on the desired level of solvation. The program outputs the expected sedimentation coefficient and both the rotational and translational diffusion coefficients, for the calculation of the hydrodynamic radius of the input model. Correlating these values to the experimental values derived from SV-AUC data fitting, relates a structural model with behaviour of that protein in solution. Using this to predict the hydrodynamic properties of the Sgt1 TPR domain shows a discrepancy in the expected molecular weight and s-value, whereby the fitted molecular weight estimate for the experimental distribution of Sgt1 1-150 does not coincide with the expected s-value for a trimer of the Sgt1 TPR domain, as determined by HYDROPRO. The model of the TPR domain used for this analysis is discussed further in Chapter 4, and encompasses the core TPR domain to residue 136.

Conversely, a dimer of Sgt1 1-136 has a predicted s-value between 2.8 to 2.9 S, depending on the level of modelled solvation. This value more closely reflects the experimental sedimentation coefficient for Sgt1 1-150 and 1-178 of 2.6 and 2.9 S, respectively, highlighting a discrepancy in the molecular weight estimate calculated from the SV-AUC data fitting for Sgt1 1-150 (assuming that the 14 amino acid difference, which is not visible in the crystal structure, does not produce a significant difference in the predicted sedimentation coefficient), and to a lesser extent 1-178.

Size exclusion chromatography-coupled multi-angle light scattering (SEC-MALS) allows the determination of an absolute mass for a species whose mass is not easily determined through the calibration of a size exclusion column, owing to

micro-heterogeneity within a sample unresolved on the time-scale of elution or because of non-spherical particle shape. Due to the discrepancy between the SV-AUC analysis and the modelled sedimentation coefficient from HYDROPRO, a SEC-MALS experiment was carried out for Sgt1 1-150 to determine whether there is micro-heterogeneity within samples of Sgt1 1-150 at different concentrations, which might influence the molecular weight estimate derived from the fit of a continuous c(s) model.

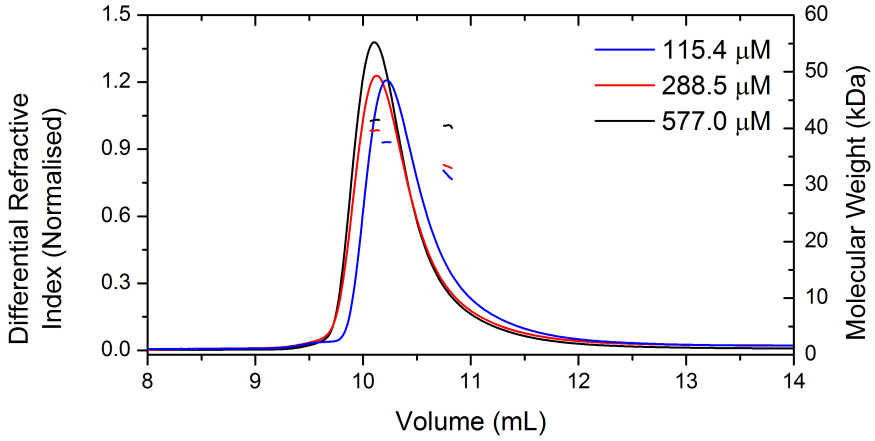


Figure 3.21: **SEC-MALS Analysis of Sgt1 1-150 –** Sgt1 1-150 was examined by SEC-MALS at different concentrations in order to identify micro-heterogeneity within the population that could have lead to discrepancies in the molecular weight estimate in the SV-AUC analysis. The centre of all peaks yields a molecular weight between a dimer and trimer of Sgt1 1-150 based on the dRI, while the trailing edge of the peak corresponds to a dimer of 1-150 at the lowest concentrations. This suggests that the single peak from SEC comprises a mixture of dimer and trimer species in solution.

Sample concentrations equivalent to 2.0, 5.0 and 10.0 mg/ml (115.4, 288.5 and 577.0 µM respectively) were applied in 100 µL volume to a Superdex 75 HR 10/30 size exclusion column pre-equilibrated in buffer containing 25 mM HEPES pH 7.0, 150 mM NaCl, 2 mM DTT, 1 mM EDTA and 1% sodium azide. The overall elution profile between concentrations did not differ, with the peak having a slight tail at the trailing edge (Figure 3.21). The single peak is centred at 10.1 mL at the highest concentration and 10.2 mL at the lowest concentration tested. A partial Zimm analysis for molecular weight determination, which takes into account the differential refractive index (dRI) and scattered light intensity of particles at a particular elution volume, suggests that the centre of all peaks

corresponds to a dimer species of Sgt1 1-150, with a molecular weight of 41452 Da at 577.0 μM and 37546 Da at the lowest starting concentration of 115.4 μM. The change in the molecular weight estimate and shift in peak centres between concentrations parallels the SV-AUC analysis, which suggested the existence of a dimer and possibly a dimer:trimer equilibrium. There is also a smaller species within the trailing edge at the two lower concentrations, with an average molecular weight of 32517 Da, which more closely approximates the expected molecular weight of a dimer of Sgt1 1-150. At the highest starting concentration, the trailing edge has a similar attributed molecular weight to the centre of the peak (40459 Da). As there is no evidence of a monomer molecular weight species, the SEC-MALS analysis supports concentration-dependent higher order oligomerisation beyond a dimeric species of Sgt1 1-150.

3.6 Analysis of Sgt1:Skp1 Complexes In Solution

Sgt1's propensity for self-association might entail many different consequences for its cellular function. In humans, Sgt1 forms a 1:1 stoichiometric complex with Skp1 (Zhang *et al*., 2008), a protein critical to the SCF ubiquitin ligase machinery, but no known function in the human kinetochore, unlike its yeast counterpart. There is currently no direct experimental evidence for the stoichiometry of yeast Sgt1 and yeast Skp1. We therefore wanted to know both how Sgt1 associates with Skp1 and what oligomeric state of Sgt1 mediates this interaction.

3.6.1 Analysis of Sgt1:Skp1 Complexes by Analytical Size Exclusion Chromatography

Analytical size exclusion experiments provide a good first estimate for association and stoichiometry of two putative partners. The elution volume is directly related to the Stokes' radius and molecular weight of the protein, which can in turn be extrapolated from the raw data. This technique was therefore used to evaluate the ability of various Sgt1 constructs to bind Skp1.

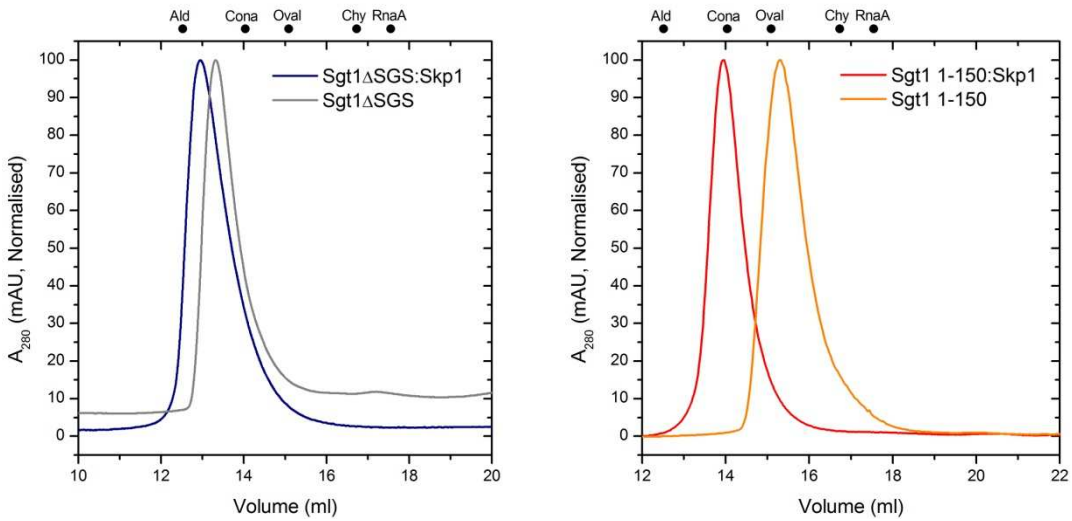
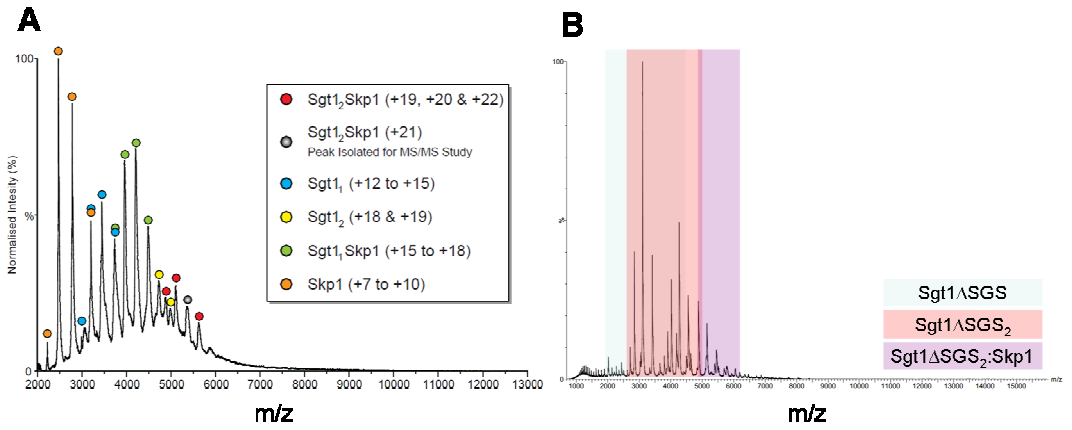


Figure 3.22: **Analytical Size Exclusion Chromatography Analysis of the Sgt1-Skp1 interaction** – These experiments were carried out using a Superdex 200 HR 10/30 column supplied by GE Healthcare at a constant flow-rate of 0.5 ml/min. The column was pre-equilibrated in buffer containing 50 mM Tris pH 7.5, 150 mM NaCl, 2 mM DTT and 1 mM EDTA. Protein samples were prepared in an identical buffer and, where necessary, mixed in a 1:1 molar ratio prior to the analysis. On the right, Sgt1ΔSGS (in **Grey**) and Skp1 complex formation (in **Blue**) is illustrated by the overlay of chromatograms and shift in elution volume. Similarly, complex formation between Sgt1 1-150 and Skp1 on the left, with Sgt1 1-150 on its own in **Orange** and the complex in **Red**. The elution volumes of molecular weight standards used in calibration of the column are indicated by ● symbols above each chromatogram.

In the absence of Skp1, Sgt1ΔSGS and Sgt1 1-150 elute at 13.3 and 15.9 ml respectively, from a column of approximately 23.6 ml (Figure 3.22, left). Through calibration of this column using standards of known molecular weight, these elution volumes can be translated to molecular weight estimates of 107800 Da for ΔSGS and 32100 Da for 1-150, both roughly equivalent to a dimer. The slightly higher molecular weight deduction for Sgt1ΔSGS likely stems from the lack of an interaction between the TPR and CS domains as well as the flexible linker separating the two, resulting in a more elongated molecule. The N-terminal construct, on the other hand, is closer to its predicted dimer molecular weight. The close estimation to the dimer molecular weight suggests that the TPR domain in this size exclusion experiment is probably globular.



Species	Stoichiometry	Expected Mass (Da)	Observed Mass (Da)	m/z Range
Sgt1 FL:Skp1	2:1	112742	115934 ± 7.36	5270.74 – 6102.80
	1:1	66856	69067.25 ± 5.70	3838.07 – 4605.48
Sgt1ΔSGS:Skp1	2:1	87384	92659.33 ± 11.07	*Data not available*
Sgt1 FL	2	91772	93734 ± 3.32	4464.52 – 5514.76
	1	45886	46867 ± 1.66	3348.64 – 3906.58
Sgt1ΔSGS	2	66414	68270.40 ± 13.19	*Data not available*
	1	33207	34133.39 ± 10.42	
Skp1 FL	2	41940	44400.5 ± 8.08	3172.46 – 4037.41
	1	20970	22200.25 ± 4.04	2221.00 – 3172.43

Figure 3.23: **Native Electrospray Ionisation Mass Spectrometry of Sgt1:Skp1 – (A)** Native ESI-MS spectrum of a complex of full length Sgt1 and full length Skp1. This preliminary spectrum shows that Sgt1 and Skp1 form a 2:1 stoichiometric complex that can be detected in the gas phase. Observed and expected masses are listed in the accompanying table. **(B)** Native ESI-MS spectrum of a complex of Sgt1ΔSGS and full length Skp1, showing the expected 2:1 stoichiometry. Observed and expected masses are listed in the accompanying table. Spectra collection and data analysis was carried out by Richard Kerr.

Complexation experiments with Skp1 were performed at 150 mM final concentration of sodium chloride. Upon the addition of Skp1, both Sgt1 peaks shift to a lower elution volume of 13.0 and 14.0 ml, with a corresponding increase in the calculated molecular weight of 126500 and 79400 Da for the ΔSGS and 1-150 complexes respectively (Figure 3.22, right). This is suggestive of complex formation with Skp1. From the primary sequence, a Skp1 monomer was expected to add an additional 22300 Da to the molecular weight of the Sgt1 dimer in isolation. While the molecular weight of the ΔSGS:Skp1 complex equates to a 2:1 stoichiometry (confirmed by ESI-MS, Figure 3.23), the 1-150:Skp1 complex is further removed from a 2:1 estimate of 57200 Da and closer to a 2:2 estimation of 79500 Da. The functional unit for binding Skp1 thus appears to be a dimer of Sgt1. In addition, the absence of the CS domain has an effect on binding of additional Skp1 proteins to the N-terminus of Sgt1. This would corroborate the previously discussed idea that sequences C-terminal to the core TPR affect the properties of Sgt1 in solution. Alternatively, because of the effect of shape on

the derivation of a molecular weight estimate, it is also possible that the observed elution volume correlates with an elongated 2:1 complex.

3.6.2 Analysis of Sgt1:Skp1 Complexes by Size Exclusion Chromatography-coupled Multi-angle Light Scattering

To confirm the stoichiometry of an Sgt1:Skp1 complex, further SEC-MALS experiments were conducted with full length Skp1 complexed to either full length Sgt1 or Sgt1 1-150.

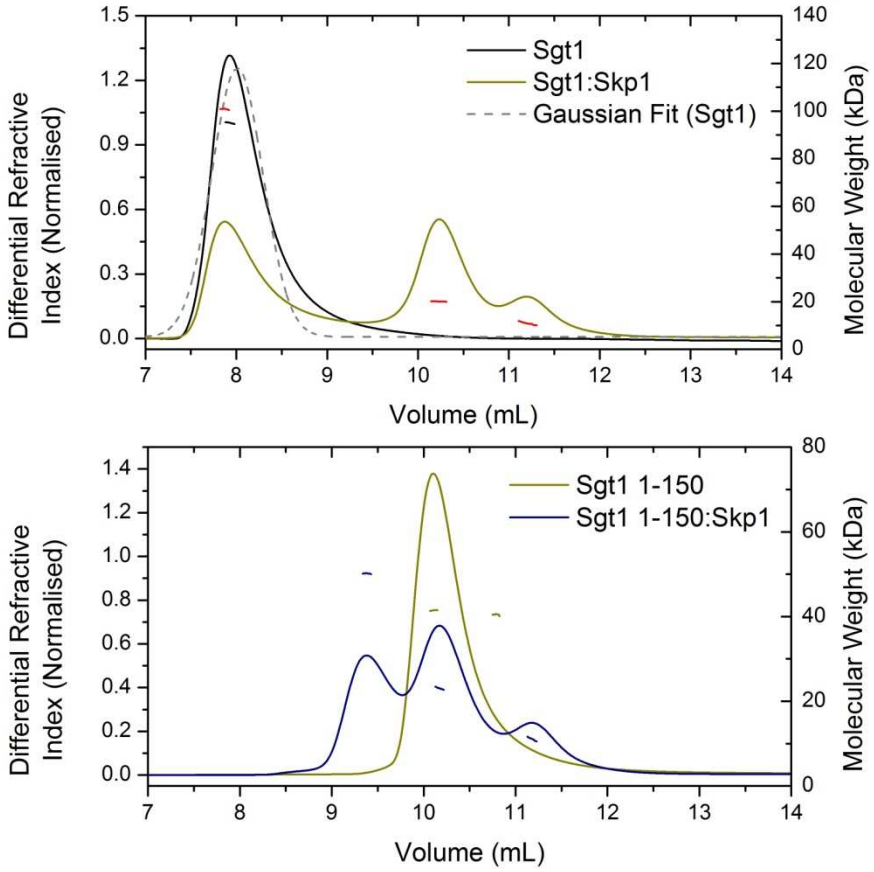


Figure 3.24: **SEC-MALS Analysis of Sgt1:Skp1 Complexes** – Based on the AUC analysis and the understanding that Sgt1 1-150 exhibits micro-heterogeneity through the co-population of dimer and trimer species, further SEC-MALS analyses were carried out to investigate the stoichiometry of Sgt1:Skp1 in solution, and to determine whether a trimer of the Sgt1 TPR domain plays a role in complex formation. Analysis of the full length proteins (**Top**) shows a slight shift in the elution volume and a small increase in the molecular weight estimate for the centre of the peaks, upon the addition of Skp1. The peak corresponding to Sgt1 1-150 on its own consists of both dimer and trimer species, with the addition of full length Skp1 producing a peak of lower elution volume corresponding to a 2:1 complex of Sgt1 1-150:Skp1 (**Bottom**). The third peak in this sample might correspond to a degradation product of Skp1.

The respective complexes were prepared with a 2-fold molar excess of Skp1 in order to promote complex formation and incubated for 24-hours prior to analysis. While only a small population of full length Sgt1 trimers were detected in the gas phase (Figure 3.9), it was necessary to account for possible conformational rearrangements upon complexation of the full length components allowing trimer formation, and ultimately confirm the expected 2:1 complex as the relevant biological entity as first suggested by SEC analysis of Sgt1ΔSGS in complex with Skp1.

At a concentration of 44.6 μM, full length Sgt1 elutes as a single peak, with the molecular weight of species eluting at the centre of the peak (at 7.9 mL and therefore within the void volume) approximating a dimer (95031 Da, with an expected monomer molecular weight 44860 Da) (Figure 3.24). A partial Zimm analysis of the scattering data yields a molecular weight estimate of 60826 Da for the trailing edge of this peak. As this molecular weight falls between monomer and dimer species, it is possible that some dissociation is occurring of the dimer. This is supported by the non-Gaussian distribution of the peak (Figure 3.20, top).

When complexed with Skp1, there is no change in the elution volume because the apparent complex elutes in the void (7.9 mL) (Figure 3.24, top). A 2:1 stoichiometric complex of full length Sgt1:Skp1 has a predicted molecular weight of 111918 Da. As the Sgt1 dimer has an elongated shape, it will have a significant influence on the molecular weight estimate as determined by light scattering. Therefore, the derived molecular weight of 100817 Da for the centre of the largest peak, which is approximately 11000 Da removed from an absolute 2:1 complex, likely reflects the expected 2:1 stoichiometry. This is exacerbated by the exchange of complex components in equilibrium during size exclusion, and the dilution effect that accompanies this process, as well as the inherent error in the molecular weight calculation (about 5% (Oliva *et al.*, 2001)). Furthermore, the molecular weight does not approximate a 1:1 or 2:2 stoichiometry (with expected molecular weights of 67059 and 134117 Da, respectively). Species eluting within the volume range of the secondary peak at 10.2 mL have an average molecular weight of 20101 Da across the fitting range. Given the 2-fold

molar excess of Skp1 in the original sample, this peak corresponds to uncomplexed Skp1.

In complexing Sgt1 1-150 with full length Skp1, it was expected that a 2:1 complex would form that would be tight enough to remain intact during size exclusion. Given that Skp1 was added in molar excess, the 10.2 mL species, with calculated mass of 23063 Da, corresponds to excess Skp1 (Figure 3.24, bottom), which has a theoretical mass of 22199 Da. Given the correlation of the calculated and estimated mass for this peak, the appearance of a new peak at 9.4 mL, which is not present in the non-complexed sample, suggests that the entire population of Sgt1 1-150 is complexed with Skp1, since the molar excess of the latter would shift all free Sgt1 in solution into the complex peak. This has an estimated molecular weight of 50159 Da, which is larger than a 1:1 stoichiometric complex (39530 Da), but slightly smaller than the theoretical mass for a 2:1 complex (56861 Da). Therefore, the relevant biological complex entails a 2:1 ratio of Sgt1 to Skp1.

3.6.3 Analysis of Sgt1:Skp1 Complexes by Sedimentation Velocity Analytical Ultracentrifugation

The SV-AUC analysis discussed in Section 3.4.2 shows that sequences between residues 151-178 of Sgt1 play a role in modulating self-association of the N-terminus and that higher order oligomers of the TPR domain may form in solution. To determine whether a trimer of Sgt1 plays a functional role in kinetochore assembly via Skp1, a titration experiment using increasing concentrations of Sgt1 1-150 was conducted to establish whether a 3:1 complex of Sgt1:Skp1 is able to form in solution (Figure 3.25).

Complexes of Sgt1 1-150 and Skp1Δ BTB/POZ were prepared in 1:1, 2:1, 5:1 and 10:1 molar ratios to a total absorbance of 1 AU, and incubated for one hour before being analysed by SV-AUC. It was determined that even at the lowest relative concentration, Sgt1 would predominantly be a trimer, as suggested by previous SV-AUC analyses of Sgt1 1-150 on its own (see Section 3.4.1). Therefore,

if a 3:1 complex can form in solution, then the analysis using a continuous c(s) distribution model would highlight a 66257 Da species when Sgt1 is in excess. A 2:1 complex would have a molecular weight of 48927 Da.

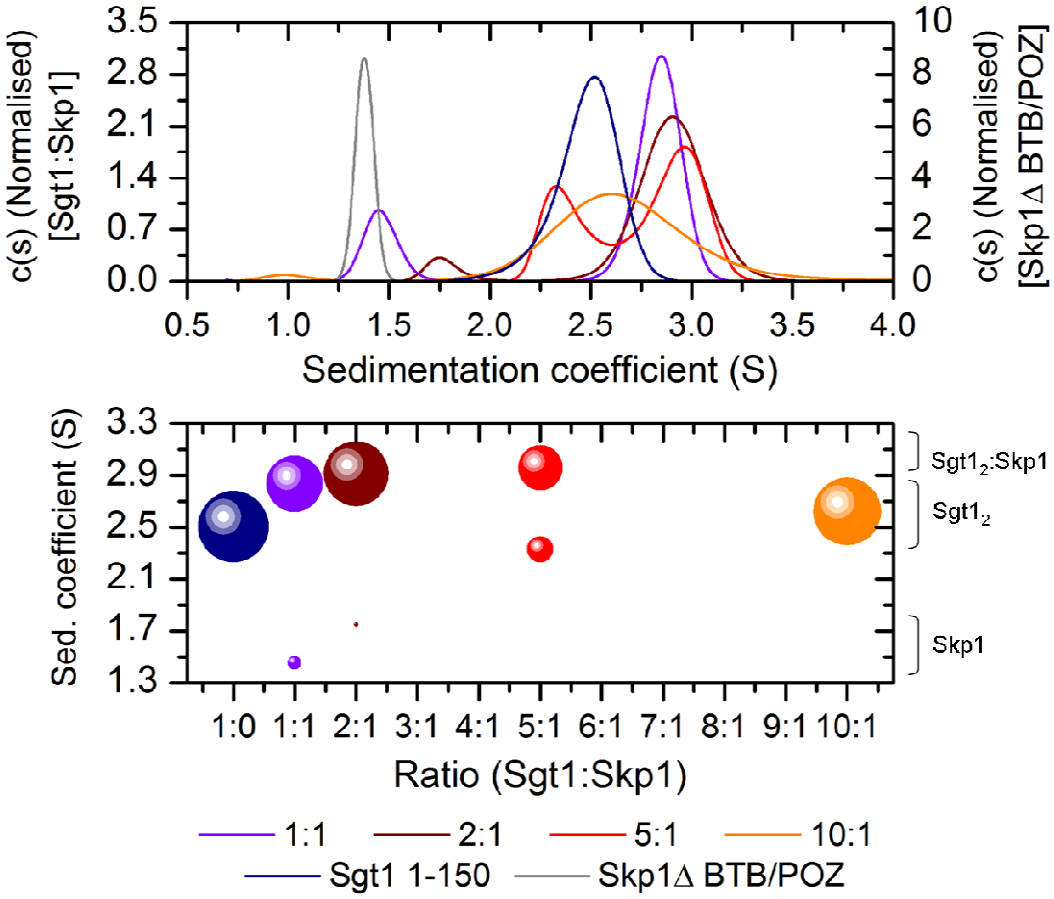


Figure 3.25: **Analysis of Sgt1:Skp1 Complex Titration by SV-AUC** – Solutions of Sgt1 1-150 and Skp1Δ BTB/POZ of known concentration were combined in varying molar ratios in order to investigate the stoichiometry of Sgt1:Skp1 complexes and probe the potential formation of a 3:1 Sgt1:Skp1 complex in solution. A 2:1 molar ratio of Sgt1 to Skp1 (**Brown**) showed a predominant peak with a molecular weight estimate of a 2:1 stoichiometric complex. 1:1 (**Violet**) and 5:1 (**Red**) ratios gave rise to two peak distributions, with the main species corresponding to a 2:1 complex of Sgt1:Skp1 as inferred by the molecular weight estimate and s-value of the main peak in the 2:1 sample. A 10:1 ratio (**Orange**) of Sgt1:Skp1 produces a single broad peak spanning s-values of both the Sgt1 dimer:trimer equilibrium and the 2:1 Sgt1:Skp1 complex. The size of the sphere corresponds to the size of the population in the sample, as derived from the total integral of each distribution.

A 2:1 ratio of Sgt1:Skp1 produces a predominant peak (90.1% of the total population) in the continuous c(s) distribution with an s-value of 2.91 (Figure 3.25, brown) and a molecular weight estimate of 50102 Da. As the minor species comprises less than 10% of the total population, the weight-average frictional

ratio derived from the data fitting process is determined largely by the main species. Therefore, the molecular weight estimate suggests that a 2:1 complex forms in solution. Similarly, at a 1:1 ratio (Figure 3.25, violet) a large 2.8 S species comprising 78.3% of the total population has a molecular weight estimate of 52173 Da, again reflective of a 2:1 complex in solution. Therefore, the 6.5% population at 1.46 S is the theoretical excess of Skp1Δ BTB/POZ, confirmed by the position of the single peak distribution corresponding to Skp1Δ BTB/POZ analysed without Sgt1 (Figure 3.25, grey).

At a 5:1 molar excess of Sgt1 1-150, two peaks are revealed by a continuous c(s) fit, that comprise 62.9% and 36.8% of the population, with s-values corresponding to a 2:1 complex (3.0 S) and excess of Sgt1 (2.3 S) respectively (Figure 3.25, red). As a 3:1 Sgt1:Skp1 complex would have caused a significant shift in the s-value of the main peak, and a decrease in a secondary population (whether Sgt1 trimer or Skp1), it is unlikely that a 3:1 complex of Sgt1:Skp1 can exist in solution at the concentrations used in this experiment. At an increased molar excess of Sgt1 to ten-fold greater than Skp1, a single species (94.9% of total population) with a broad distribution centred at a lower s-value of 2.6, that is similar to previous analysis of Sgt1 1-150 (Figure 3.16), represents an equilibrium that is predominantly Sgt1 1-150 (Figure 3.25, orange).

A small shoulder at the trailing edge of the peak suggests that some complex formation is occurring, though if a 3:1 complex was being formed, then this shoulder would likely be positioned differently to reflect a larger s-value and peak centre. Therefore, even at a significant molar excess and at relative concentrations at which Sgt1 is definitively a trimer, a larger species with a higher s-value does not form.

Trimers of Sgt1 therefore do not form complexes with Skp1 at the concentrations examined by SV-AUC, even when Sgt1 is in molar excess, which infers that the functionally relevant stoichiometry of the Sgt1:Skp1 interaction is 2:1, and that only one dimer conformation of Sgt1 (of the two inferred by analysis of Sgt1 1-178) supports complex formation. These observations were confirmed by mass

spectrometry experiments, which showed that an *in vitro* reconstituted 2:1 complex of Sgt1ΔSGS and full length Skp1 (and of Sgt1 1-150 and Skp1, data not shown) is tightly associated and a 3:1 complex does not form (Figure 3.21).

CHAPTER 4: STRUCTURAL ANALYSIS OF AN SGT1:SKP1 COMPLEX AND THE SGT1 TPR DOMAIN

The understanding of Sgt1's role in kinetochore assembly is currently centred around two ideas: firstly, that Sgt1 dimerisation is required for kinetochore assembly (Bansal *et al*., 2008), and secondly, that this process is reliant on an interaction with the CBF3 component Skp1 (Catlett *et al*., 2006). Bansal *et al.* (2008) showed that dimerisation of Sgt1 was important for its kinetochore-specific function and that disrupting dimerisation by mutating a leucine – positioned in the core TPR fold – to a proline abolished Skp1 binding. Whether or not this leucine is a residue essential to the fold of a TPR repeat, their experiments highlight the importance of the TPR domain of Sgt1 for its function. Previously discussed biophysical studies of the Sgt1 TPR domain have directly confirmed that the TPR domain is sufficient for self-association and that this domain also binds Skp1 (see Chapter 3). While TPR domains are known to self-associate, structural predictions and sequence analyses of Sgt1 (to identify features conducive to self-association) do not provide concrete evidence for its mode of dimerisation. A structure of Sgt1 would therefore provide structural insight into its mode of dimerisation, while the structure of a complex of Sgt1 and Skp1 would shed light on the functionally relevant dimer for kinetochore formation.

4.1 CRYSTALLISATION OF AN SGT1:SKP1 COMPLEX

With knowledge of the solution behaviour of Sgt1 and the assumption that the stoichiometry of Sgt1:Skp1 in solution is 2:1 (see Chapter 3), preparation of the complexes for crystallisation was performed *in vitro* using purified protein constructs described in Section 3.2, with the protocol outlined in *Chapter 2*. Sgt1 1-150 or 1-178 was pooled with Skp1Δ BTB/POZ protein at a 2:1 molar ratio and incubated for one hour (Figure 4.1). The final concentration used for crystallisation was 25.5 mg/ml.

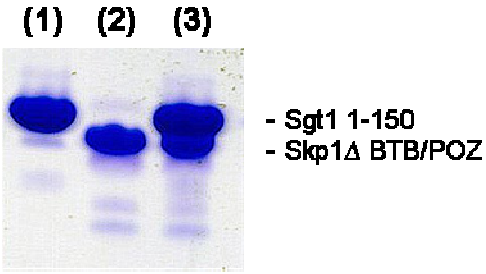


Figure 4.1: **SDS-PAGE Analysis of an *in vitro* Reconstituted Sgt1:Skp1 Complex** – Purified Sgt1 and Skp1 constructs were pooled together in a 2:1 (Sgt1:Skp1) molar ratio for crystallisation. The adjacent SDS-PAGE analysis illustrates the composition of this reconstituted complex, with purified Sgt1 1-150 in **Lane (1)**, purified Skp1 (at half the molar concentration of Sgt1) in **Lane (2)** and the complex in L**ane (3)**.

An initial screen of crystallisation conditions yielded four hits for a complex of Sgt1 1-150:Skp1Δ BTB/POZ with similar crystal morphologies after approximately 30 days of incubation at 4°C (Figure 4.2). Crystals grew in both 1:1 and 2:1 ratio drops of protein to reservoir solution. The crystals had a needle-like shape with rough edges, with the higher protein:reservoir ratio yielding large single crystals that could be harvested and tested for diffraction. UV screening of the crystals via a Rigaku plate imaging system confirmed that these were protein crystals. A single condition was chosen for a simple 96-well grid screen optimisation (JBScreen PACT++ D10.2), testing the effects of $MgCl_2$ and precipitant (PEG 6000) on crystal formation.

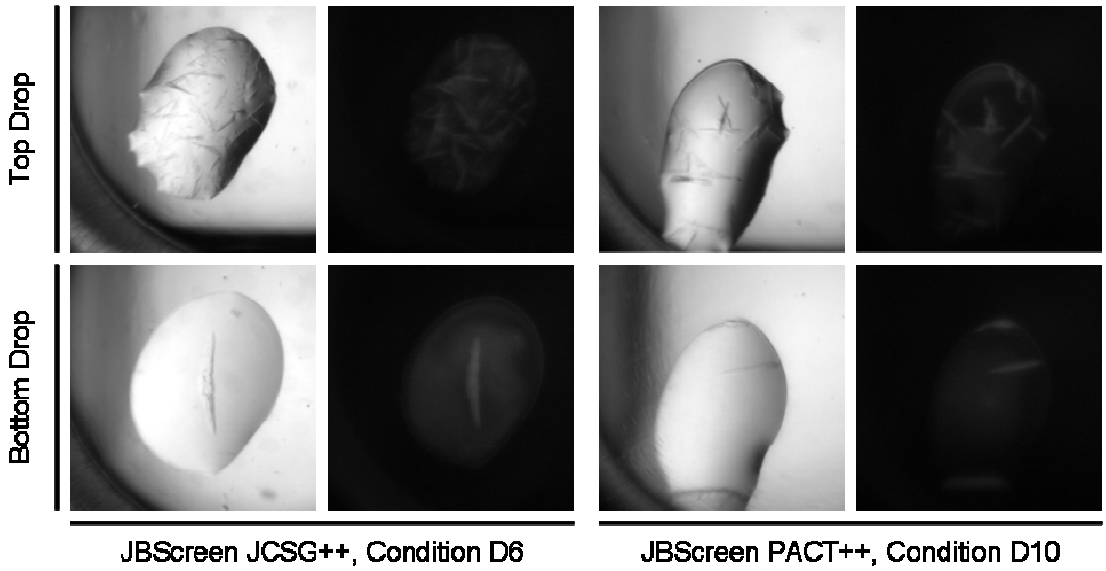


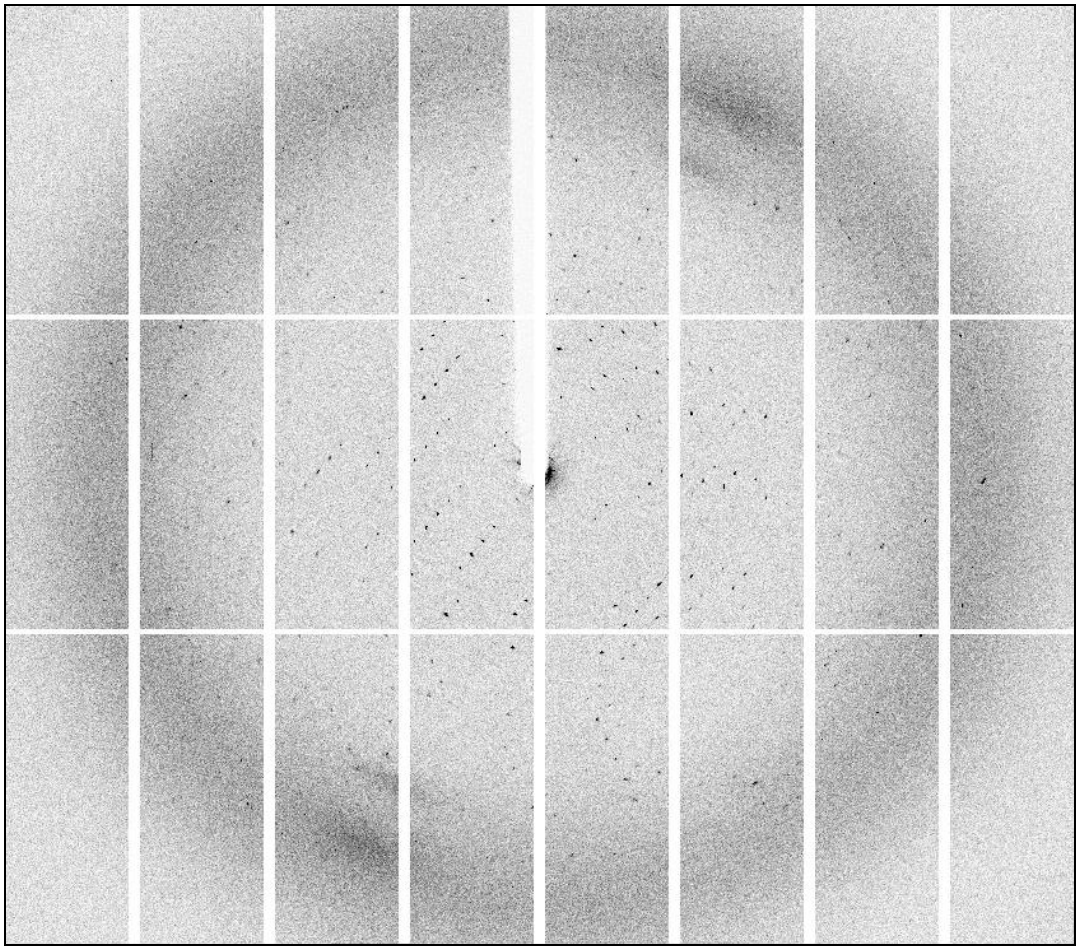
Figure 4.2: **Initial Crystals of Sgt1 1-150: Skp1Δ BTB/POZ Complex** – A screen of 384 conditions gave crystals in both drops of two conditions with similar buffer components. The original screen was set up with a 1:1 ratio of protein to reservoir solution in the top drop of a 96-well Maxwell plate, with a 2:1 ratio in the bottom drop. Total drop volume was 150 nL. All crystals were of similar morphology, with the bottom drop yielding bigger, single crystals, while the top drop gave rise to clusters of crystals. Crystal formation took approximately four weeks. **(JCSG D6)** 20% PEG 8000, 100 mM Tris-HCl pH 8.0, 200 mM Magnesium Chloride. **(PACT D10)** 20% PEG 6000, 100 mM Tris-HCl pH 8.0, 200 mM Magnesium Chloride.

The subsequent optimisation screens gave crystals of several different morphologies in different drops: feather-like crystals that appeared in clusters throughout the crystallisation plate, needle-like crystals similar in overall morphology to the original hit, but markedly thinner across (data not shown) and cube-shaped crystals, seen in crystallisation attempts of Sgt1 1-150 (see Section 4.9).

4.2 Attempts at Solving the Structure of an Sgt1:Skp1 Complex by Molecular Replacement

Needle-like crystals of a complex of Sgt1 and Skp1 were grown in, 100 mM Tris-HCl pH 8.0, 200 mM $MgCl_2$ and 20% (w/v) PEG 6000. Crystals were frozen in cryoprotectant solution comprising the reservoir solution with 20% (w/v) ethylene glycol. Diffraction experiments were conducted on IO4-1 at the Diamond Light Source synchtrotron. Based on the input strategy from an initial scan, 650 images were collected giving a final multiplicity of 7. Data was processed locally at the synchrotron via the XDS pipeline (Kabsch, 2009) and utilised for molecular replacement attempts in Phaser (McCoy *et al*., 2007), Molrep (Vagin *et al*., 1997) and MrBUMP (Keegan *et al*., 2007).

Despite low sequence identity, the core topology of a TPR motif is conserved amongst all known TPR proteins. As a direct consequence of this, at least 27 deposited structures of TPR domains have been solved by molecular replacement using existing structures, including the structure of the two- and three-repeat consensus TPR domains solved by Regan and colleagues, who utilised a polyserine ensemble of existing TPRs to develop a suitable molecular replacement model (Main *et al*., 2003). Despite the fact that Sgt1 has low sequence identity to other TPR domains, a polyalanine or polyserine model might provide enough phasing power for structure determination due to the conservation of the tertiary arrangement of TPR domains.



	Overall	Inner	Outer			Overall	Inner	Outer
High resolution limit	3.27	14.63	3.27		Anomalous multiplicity	n/a		
Low resolution limit	69.29	69.29	3.36		Anomalous correlation	n/a		
Completeness	99.5	99.2	99.9		Anom. compl.	n/a		
Multiplicity	7.0	6.0	7.3		Anom. slope	n/a		
I/sigma	16.4	55.8	2.5		Partial bias	0.000	0.000	0.000
Rmerge	0.106	0.023	0.655		Total obs.	74721	938	5735
Rmeas (I)	0.123	0.026	0.655		Total unique	10659	156	784
Rmeas (I±)	0.121	0.026	0.750		Rpim (I)	0.044	0.010	0.262
Wilson B factor	88.189				Rpim (I±)	0.027	0.011	0.058

Unit Cell Dimensions	a (Å)	96.91	α (°)	90		Space Group	P 3_1 2 1
	b (Å)	96.91	β (°)	90			
	c (Å)	122.82	γ (°)	120			

Figure 4.3: **Diffraction Image and Statistics for a 3.27 Å Native Dataset from a Crystal of an Sgt1:Skp1 Complex**

As a prototypical three-repeat TPR domain, CTPR3 (Main *et al*., 2003) has significant predicted structural homology to Sgt1, and has a 2.15, 2.57 and 1.66 RMSD from the TPR domains of Hop (TPR1 and TPR2A) and PP5 respectively, which are also three-repeat TPRs. The relatively low RMSD underlines the suitability of molecular replacement to structure solution of the Sgt1 TPR

domain. In addition, given the stoichiometric excess of Sgt1 in complex with Skp1, the largest portion of the scattering mass would originate from Sgt1. The remaining phase information could then be derived from a secondary search for the BTB/POZ domain of Skp1, with the side chains in place for additional phasing power.

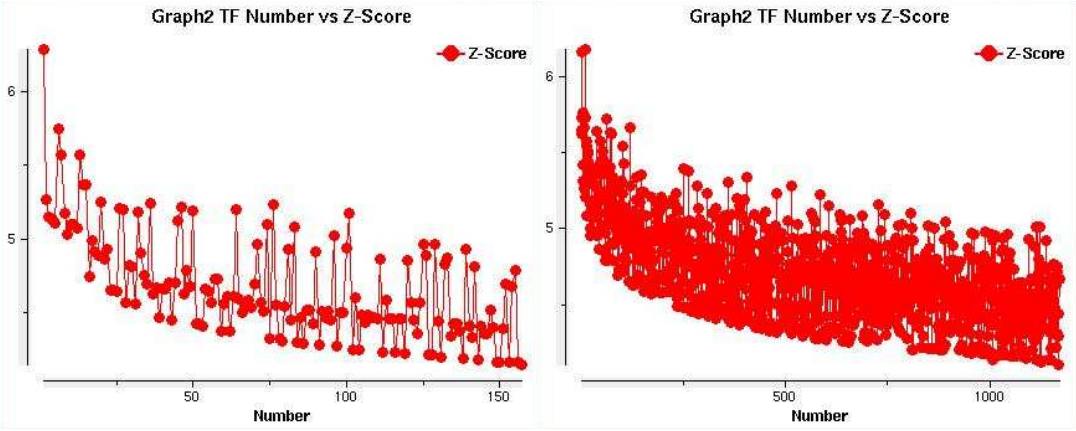


Figure 4.4: **Typical Translation Function Z-Scores from Molecular Replacement Attempts on Initial Sgt1:Skp1 Complex Data Set Using Phaser** – Searches for molecular replacement solutions in Phaser yielded low Z-scores and no increase in LLG with each step in the search. Typically, while the first search produced between 10-200 solutions **(Left)**, the subsequent search for additional molecules in the asymmetric unit gave thousands of potential solutions with low Z-scores **(Right)**.

Initial attempts revolved around the use of Phaser (McCoy *et al.*, 2007) to find a molecular replacement solution using various models of the consensus TPR and other TPR domains. Several different models were utilised for molecular replacement:

- A TPR ensemble of domains with greater than 20% sequence identity to Sgt1.
- Polyalanine models of various TPR monomers and dimer models.
- Models of the Sgt1 TPR domain generated by SWISS-MODEL (Arnold *et al.*, 2006; Kiefer *et al.*, 2009; Peitsch, 1995).
- Various models of Skp1, including full length Skp1 and the BTB/POZ domain.
- Single TPR repeats from the consensus three-repeat TPR structure (CTPR3).

Each of these models was used in searches for one, two, three and four copies of Sgt1 in Phaser, including a search for two Skp1 molecules in the asymmetric unit. TPR models were also truncated at the capping helix or within loop regions, under the assumption that these might be fundamentally different in Sgt1. Each iterative Phaser search produced Z-scores from 4.0 to 5.0 in the rotation function search, with log-likelihood gains (LLG) typically below 30, indicating an insignificant search result. A similar trend was seen for translation function Z-scores and LLG, with no increase in Z-score or LLG with each successive search for further molecules in the asymmetric unit (Figure 4.4).

Various applications of MrBUMP (Keegan *et al.*, 2007), in conjunction with Molrep (Vagin *et al.*, 1997), produced a single result with a CONTRAST score of 13.5, employing an exhaustive search of polyalanine models of various TPR domains. However, upon inspection of the electron density, it was found that there was no continuous density into which to re-build and model the Sgt1 TPR domain.

4.3 Structure Solution of the Sgt1:Skp1 Complex by Single-wavelength Anomalous Dispersion (SAD) using Selenomethionine-labelled Sgt1

Selenomethionine-labelling is a technique for the experimental derivation of phase information, which takes advantage of a tuneable X-ray source in order to derive starting phase information through a single-wavelength (SAD) or multi-wavelength anomalous dispersion (MAD) experiment.

It is now possible to utilise any bacterial cell type for expression, by employing a method of feedback inhibition and selenomethionine supplementation upon induction in order to assist the incorporation of the selenium heavy atom into the recombinant protein. The LysY/I^Q cell strain could therefore be used to maintain high levels of Sgt1 1-150 expression. Sgt1 1-150 contains three methionines, including the N-terminal methionine that was previously shown by

mass spectrometry to be clipped during recombinant protein expression (data not shown). Selenomethionine substitution for the two remaining methionines was confirmed by denaturing mass spectrometry and was highly efficient (Figure 4.5). A single selenomethionine increases the mass of unlabelled Sgt1 1-150 by 46.896 Da (93.79 Da for two). Therefore, it was determined that the majority of the purified product contained two selenomethionines, with a small amount of both singly and unlabelled protein (Figure 4.5).

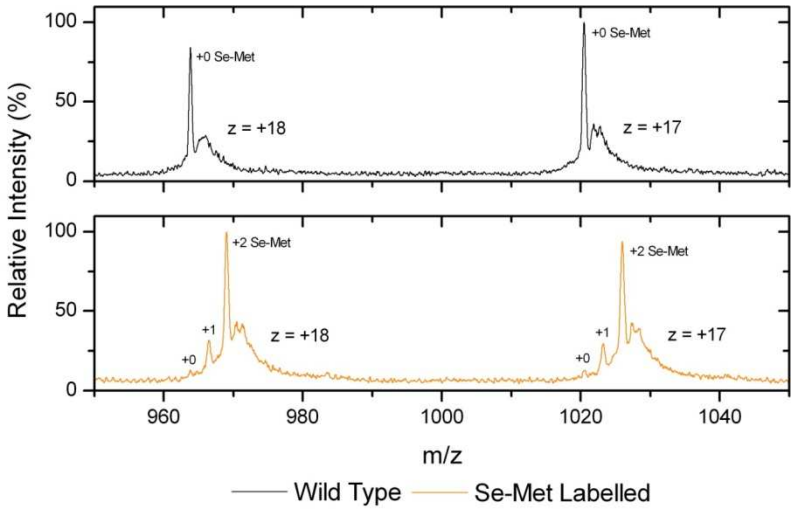
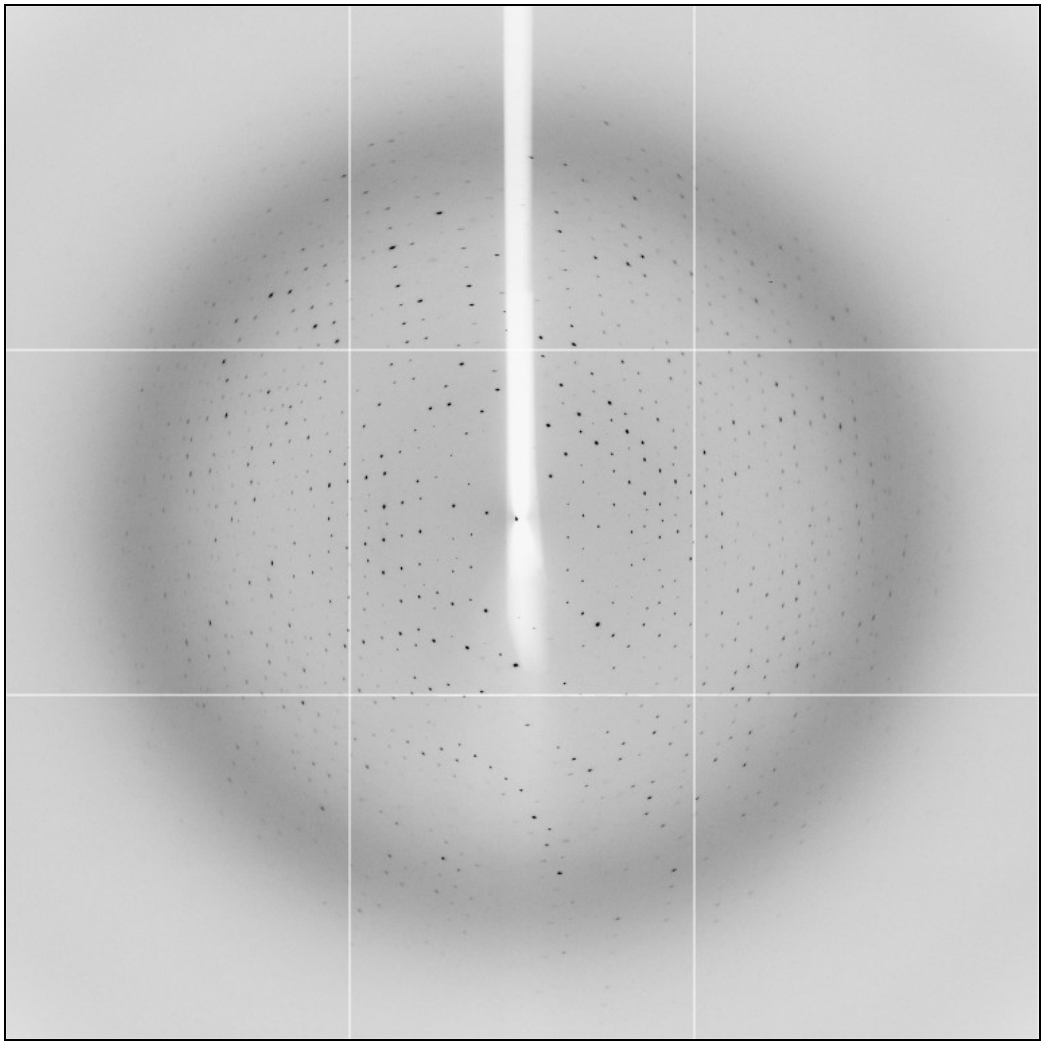


Figure 4.5: **Denaturing Mass Spectrometry Analysis of Selenomethionine-labelled Sgt1 1-150 –** Selenomethionine-labelled Sgt1 1-150 was used to facilitate initial phase derivation. Both the wild type and labelled protein were denatured in 50% acetonitrile and analysed by electrospray ionisation mass spectrometry. The spectra show successful two-fold incorporation of selenomethionine into the recombinant protein.

Crystals of selenomethionine-labelled Sgt1 in complex with unlabelled Skp1 were grown in 100 mM Tris-HCl pH 8, 325 mM $MgCl_2$ and 22.5% PEG 6000. These were harvested in reservoir solution supplemented with ethylene glycol as a cryoprotectant to a final concentration of 20% (w/v). A SAD dataset was collected of a crystal diffracting to 2.8 Å on beamline IO4 at the Diamond synchrotron X-ray source after first determining the absorption edge via a fluorescence scan. An initial inspection of this data set showed that the diffraction pattern has singular spots with no anisotropy (Figure 4.6, top). Data processing in XDS determined 100% completeness for anomalous data and a high overall anomalous multiplicity of 11.5 (Figure 4.6, bottom).



	Overall	Inner	Outer		Overall	Inner	Outer
High resolution limit	2.82	12.61	2.82	Anomalous multiplicity	11.5	10.7	11.4
Low resolution limit	81.89	81.89	2.89	Anomalous correlation	0.698	0.908	0.039
Completeness	100.0	98.3	100.0	Anom. compl.	100.0	100.0	99.9
Multiplicity	21.5	16.1	21.9	Anom. slope	1.438	0.0	0.0
I/sigma	29.2	60.8	4.9	Partial bias	0.0	0.0	0.0
Rmerge	0.089	0.034	0.826	Total obs.	339894	3444	24913
Rmeas (I)	0.099	0.051	0.863	Total unique	15826	214	1140
Rmeas (I±)	0.093	0.036	0.865	Rpim (I)	0.021	0.012	0.184
Wilson B factor	80.780			Rpim (I±)	0.027	0.011	0.255

Unit Cell Dimensions	a (Å)	94.56	α (°)	90	Space Group	P 3_1 2 1	
	b (Å)	94.56	β (°)	90			
	c (Å)	122.99	γ (°)	120			

Figure 4.6: **Diffraction Image and Statistics for a 2.82 Å Single-wavelength Anomalous Dispersion (SAD) Dataset from a Crystal of an Sgt1:Skp1 Complex**

The structure of Sgt1 1-150 in complex with Skp1Δ BTB/POZ was solved using a combination of automated phasing utilities to obtain starting phase information,

several rounds of manual building and refinement, and a single instance of molecular replacement for the Skp1 component.

The 2.8 Å SAD data set was first processed via the Auto-Rickshaw server (Panjikar *et* al., 2005) with the aim of obtaining starting phase information and a partial map to begin manual building. This server, hosted at the EMBL in Hamburg, provides an automated method for obtaining starting phases from data obtained at a synchrotron source via its advanced mode. The server employs SHELXD and SHELXE (Sheldrick, 2010) to determine the position of heavy atoms and handedness of the data set, RESOLVE (Terwilliger, 2000) or DM (Cowtan, 1994; Winn *et al*., 2011) for density modification and phase improvement, and BUCCANEER (Cowtan, 2006) to initiate model building if both of these steps are successful. By feeding in the amino acid sequence of a heavy atom-labelled protein, or providing a suitable molecular replacement solution, the Auto-Rickshaw server can provide a significant head start in model building.

Cycle	R_{Work} (%)	R_{Free} (%)	Figure of Merit [FOM]
1	41.11	44.44	0.711
2	35.97	41.94	0.726
3	33.89	40.70	0.737
4	32.95	39.51	0.743
5	32.58	39.82	0.744
6	32.36	39.30	0.746
7	31.72	39.02	0.748
8	31.57	38.98	0.749
9	31.54	39.02	0.749
10	31.51	39.07	0.749
11	31.49	38.99	0.749

Table 4.1: **R-factor Progression Throughout Buccaneer Build Cycles of the Sgt1:Skp1 Complex**

Substructure determination and initial phase calculation was conducted at a maximum resolution of 3.3 Å in SHELXD, finding 7 heavy atom signals. Refinement of heavy atom occupancy was conducted in MLPHARE (Winn *et al*., 2011), confirming 6 of 7 heavy atom sites and the opposite handedness to the spacegroup of the input data set (i.e. P 3_2 2 1) as determined by SHELXE. Density modification and phase extension to 2.8 Å in DM, with a final figure of merit of 0.639, in turn provided sufficient phase information to construct a partial alpha-

helical model in HELICAP (Morris *et al.*, 2004). No NCS averaging was conducted at this stage.

506 residues were built into the electron density from the HELICAP model by BUCCANEER, covering approximately 88.9% of the TPR scattering mass in the asymmetric unit. *Table 4.1* shows the progression of R-factors from the first cycle of model building to the final R_{Free} and R_{Work} of 38.99 and 31.49% respectively for the starting model.

After building most of the backbone of TPR-like molecules in the asymmetric unit, the Skp1 BTB/POZ domain was fitted into the map by molecular replacement in Phaser, using a model of the BTB/POZ domain constructed from PDB ID 3MKS. Phase information from the built TPR component was used to facilitate the MR search.

Subsequently, the model was further built in Coot and refined in Buster (Bricogne, 1993), using non-crystallographic symmetry restraints (for the Sgt1 protomers in the asymmetric unit) and TLS parameters derived from the TLSbasic macro. The model was refined in Buster to a final R_{Free} of 23.70% and R_{Work} of 20.22%, with 25 waters placed around protein chains in the final structure.

A Ramachandran analysis of the overall structure (Figure 4.7) shows a favourable geometry for 99.19% of all amino acid residues within the secondary structure of the asymmetric unit, with only 5.12 steric overlaps per 1000 atoms, 0.32% poor rotamers and no Cβ deviations throughout the model, as suggested by geometric analysis using the Molprobity server (Chen VB *et al.*, 2010). The minor clash score is likely due to the incompleteness of modelled side chains, which might otherwise result in steric clashes. The high solvent content of 60.8% and minimal contacts of various components of the asymmetric unit infers that many of the exposed side chains have a large degree of conformational freedom. Nonetheless, the final model is placed in the 100th percentile of Molprobity scores, with a score of 1.27.

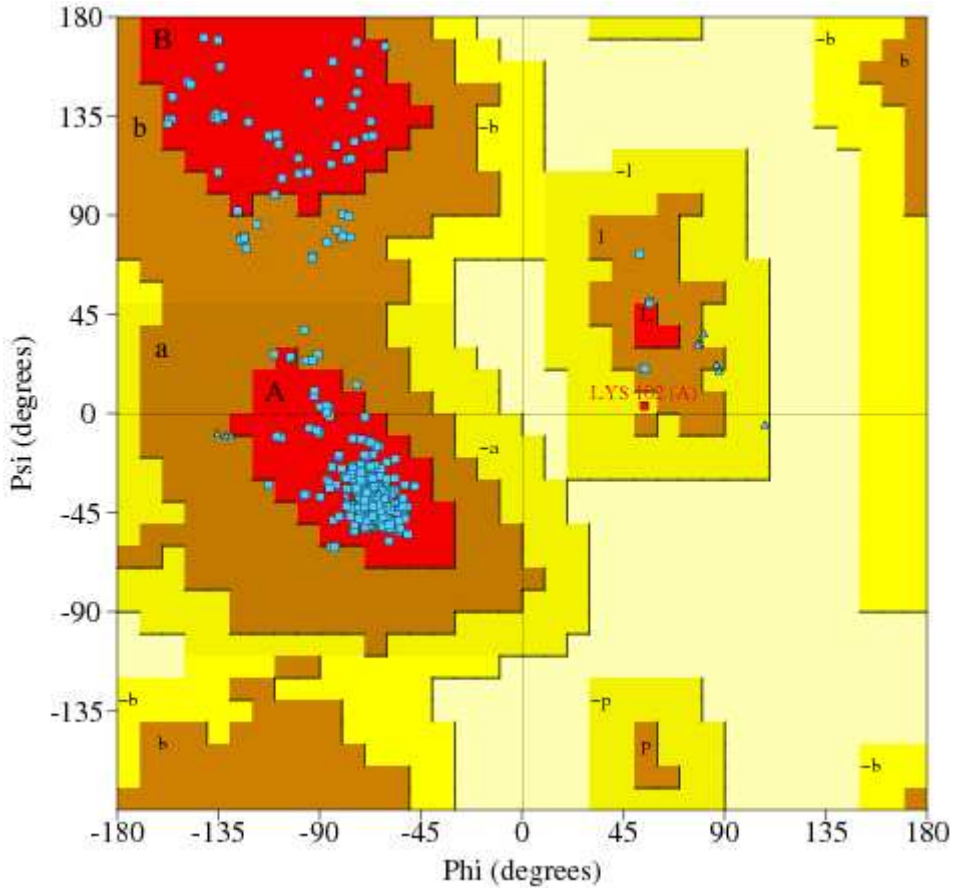


Figure 4.7: **Ramachandran Plot for the Structure of the Sgt1:Skp1 Complex –** After final refinement, the geometry of the structure was checked using the online server Molprobity (Chen VB *et al*., 2010). 99.19% of all residues were in favourable regions in the final structure refined to an R_{Work} of 20.22% and R_{Free} of 23.70%. The final Molprobity score was 1.27. This figure was generated using PDBsum (Laskowski, 2009).

4.4 Composition of the Asymmetric Unit of an Sgt1:Skp1 Complex Solved at 2.8 Å Resolution

The solved structure of Sgt1 1-150 in complex with Skp1Δ BTB/POZ comprises three copies of Sgt1 (chains A, B and C) in the asymmetric unit with a single Skp1 BTB/POZ domain (chain D) bound within the concave groove of TPR protomer chain C (Figure 4.8). Each TPR domain comprises seven alpha helices arranged as anti-parallel pairs, which together encompass three loop-separated TPR repeats and a C-terminal capping helix. The packing of the three TPRs forms the basis of the superhelical twist evident in each TPR protomer. The overall arrangement of molecules in the asymmetric unit confers a twisted V-shape to the complex.

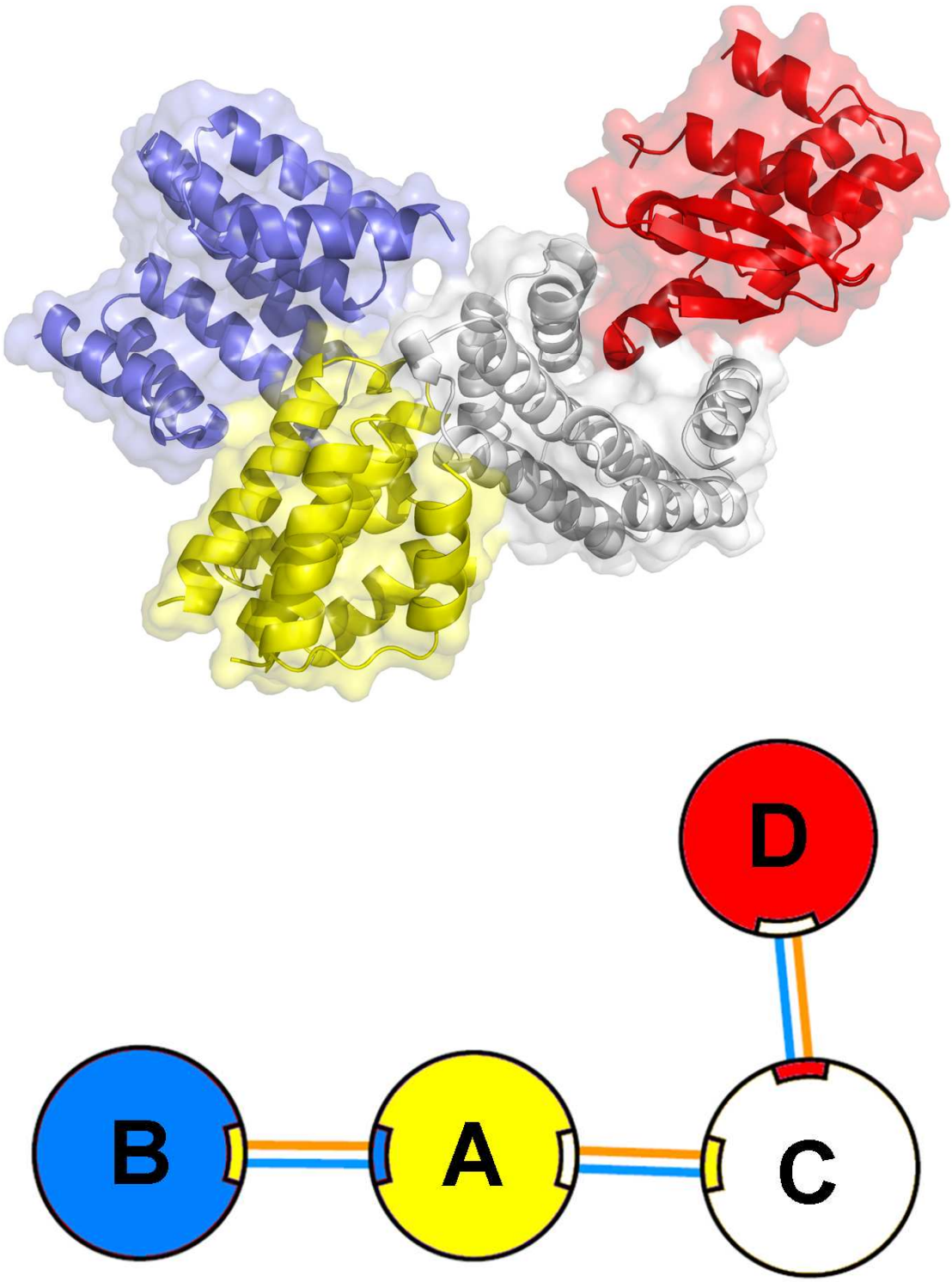


Figure 4.8: **Cartoon Representation of the Structure of the Sgt1:Skp1 Complex** – The structure of the Sgt1:Skp1 complex shows a 3:1 arrangement of Sgt1 to Skp1 within the asymmetric unit, with Skp1 (**Red**) binding in the concave groove of a single TPR protomer (**White**). The self-association interface of Sgt1 is identical between the white and yellow, and yellow and blue chains, involving interactions between the loop insertions of adjacent monomers. The domain diagram (**Bottom**) was generated using PDBsum (Laskowski, 2009).

The arrangement of TPR protomers in the asymmetric unit is a back-to-front (concave-to-convex) configuration (Figure 4.8) The map provides significant density to residue Asn136 in the most complete TPR protomer chain C, with well defined density for the protein backbone within the loop regions between adjacent TPR helices (Figure 4.9, red arrows). Chain C contacts Skp1, chain A and a symmetry-related chain B. This latter interaction occurs via the inter-TPR loops, forming a loop-mediated TPR:TPR stacking interaction with an extended concave face (see later discussion). The two molecules forming these contacts (chain C:chain B_{Sym}) are related by a two-fold symmetry axis with a total interface area of 763.4 $Å^2$. It is these extensive contacts within the asymmetric unit and with other crystallographically-related molecules that account for the completeness of the electron density for this molecule (chain C). A further discussion of the specific residues involved at these interfaces follows in Section 4.7.

In contrast, the absence of significant surface contacts in the loop regions of chain A results in discontinuous electron density (Figure 4.9, yellow). Chain A is contacted on both its convex and concave face by other TPR protomers in the asymmetric unit, but is not stabilised by interactions within its loop regions. Chain B, being positioned at one extreme end of the 3:1 crystal complex, also interfaces via the aforementioned interaction with a symmetry-related chain C, and has intermediate continuity in its electron density compared to chains A and C.

The capping helix, which is thought to be crucial for the solubility of TPRs *in vitro* (D'Andrea *et al.*, 2003), is positioned identically in each of the three TPR monomers, packing against the 3B helix of the core TPR motif. The backbone is largely resolved for this region of the TPR, with the exception of chain A, where there is weak density from Trp127 to Glu132 (Figure 4.10). One critical side chain discrepancy occurs at Trp127, one of only two large aromatic residues in the TPR domain (Figure 4.11). This residue is physically centred in a hydrophobic patch at the C-terminus of the TPR, but is differentially resolved between the different TPR chains. As with the main chain density for the core

TPR domain, the extent of side chain density in the capping helix is a direct result of the interaction of chain C with Skp1, for which Trp127 is critical (see Section 4.6).

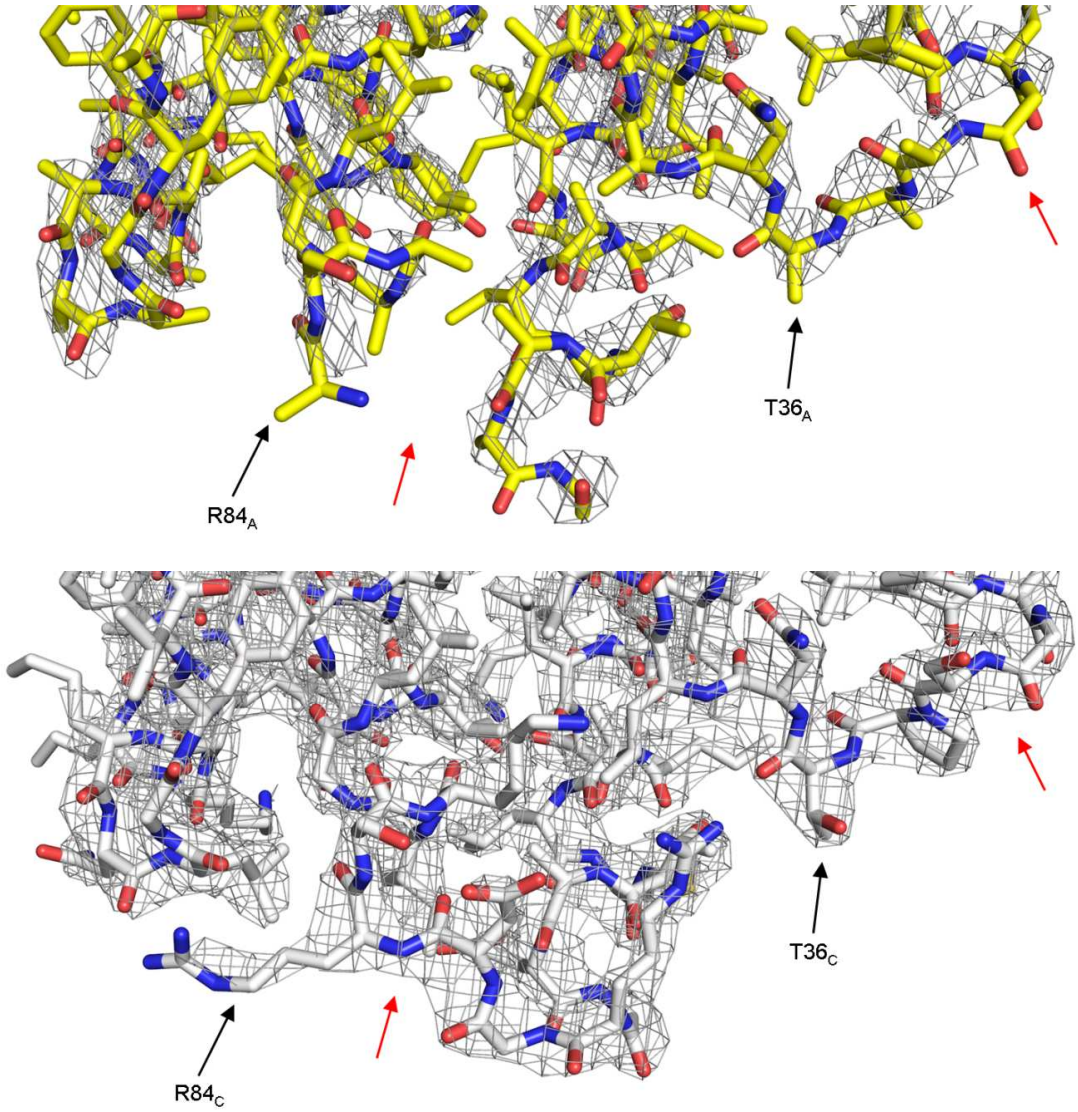


Figure 4.9: **Comparison of Electron Density Between Chains of Sgt1 in the Structure of the Sgt1:Skp1 Complex –** The differences in modelling of each chain is most easily explained by the discrepancies in continuous $2F_O$-F_C density contoured at 1.48 sigma. Comparing Chain A (**Top**, **Yellow**) to Chain C (**Bottom**, **White**) showcases the differences in continuous density, in particular within the loop regions of the TPR (**Red Arrows**). **Black Arrows** highlight differences in side chain density between Chain A and Chain C at annotated sites, with chain IDs in subscript.

The map was sufficient for building three intact TPR domains. However, there was no density corresponding to residues 137-150 of Sgt1 for any chain, suggesting that these sequences do not form any stable secondary structure and are not stabilised by components within the crystal.

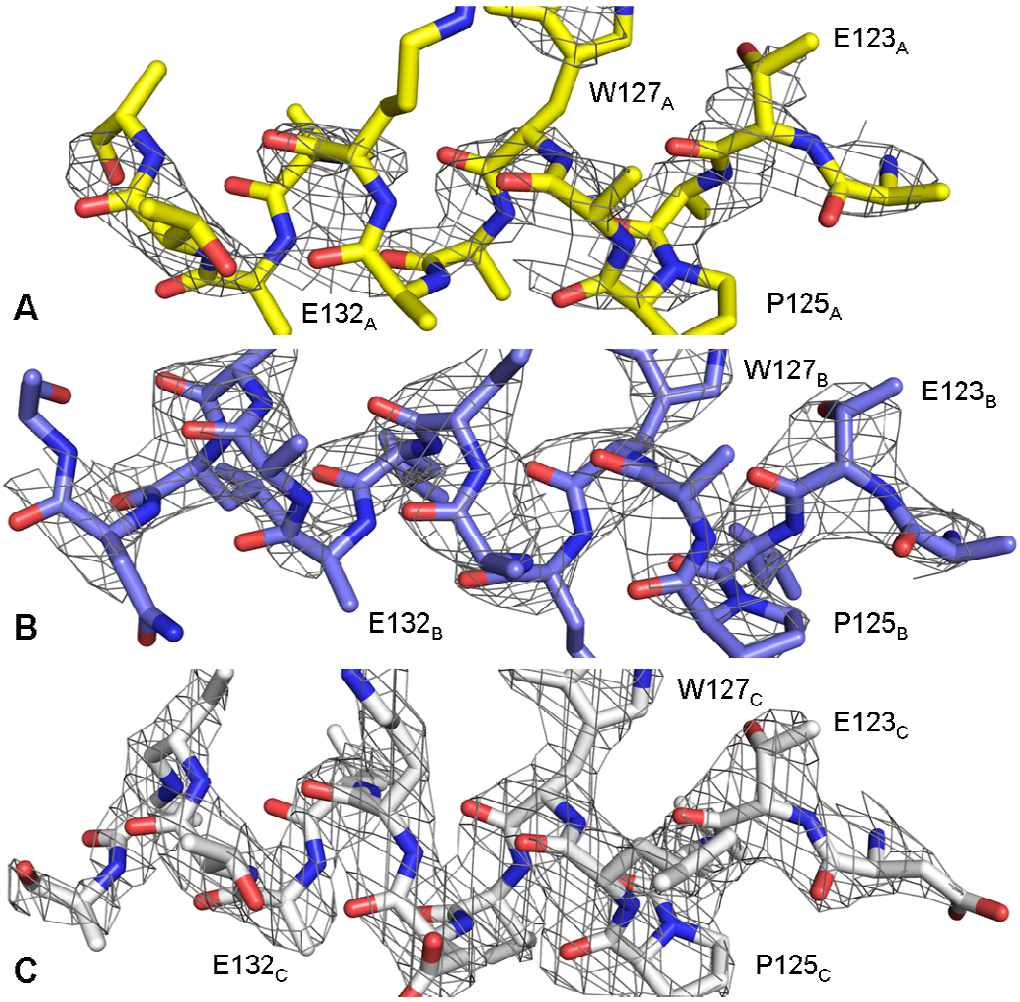


Figure 4.10: **Evaluation of Capping Helix Electron Density Between TPR Protomers in the Asymmetric Unit** – The $2F_O$-F_C electron density of the capping helix in the structure of the Sgt1 TPR domain is a good indication of how well it is stabilised by interactions with other molecules in the crystals, with the most significant density visible for the capping helix of chain C (**White, C**). The electron density here is contoured at 1.48 sigma.

The density corresponding to the BTB/POZ domain of Skp1 is weak in the areas making little contact with Sgt1. The loop region between residues 101-122 makes no contacts with other molecules in the crystal such that it remains unresolved between residues Asn101 and Asp117. This loop is also typically disordered in most other crystal structures of Skp1. Similarly the loop region between residues 35-74 is only resolved to Ser37 at the N-terminal and from Glu74 at the C-terminal end. Serine 37 forms part of the linker replacing this loop in the crystallised BTB/POZ construct. The domain interacts with the concave face of a single TPR in the ASU (chain C) and makes an additional contact on its opposite side with a symmetry-related TPR domain. Even though crystallisation was attempted in the presence of reducing agent, evidence for disulphide bond

formation between two symmetry related Skp1 molecules came to light during refinement. Cysteine 151 – which is solvent exposed in the absence of a binding partner – was seen to reside in a region of positive F_O-F_C density in the intermediate stages of refinement, with the Cα atoms separated by a distance of 6.4 Å (Figure 4.12).

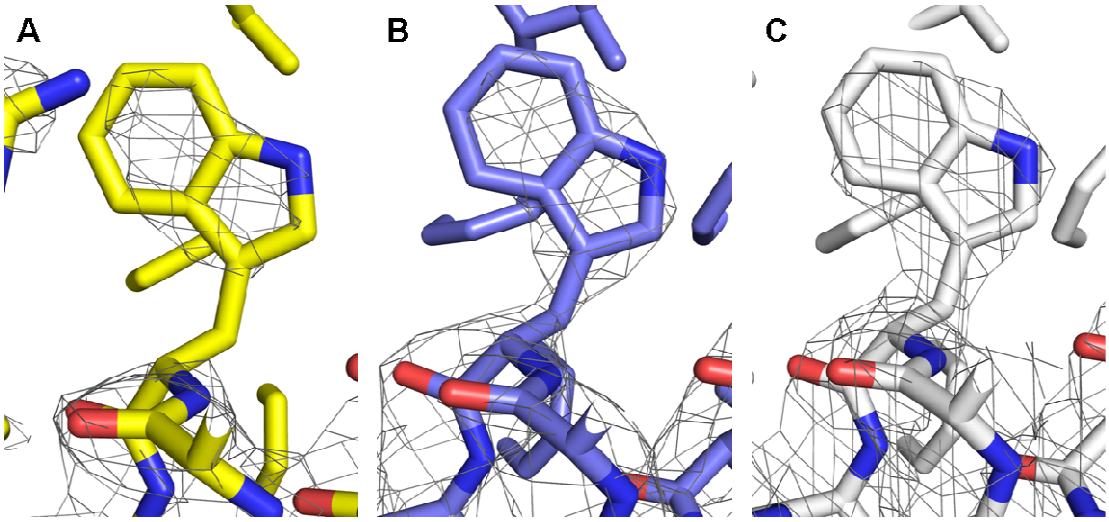


Figure 4.11 **Density Comparison of Tryptophan 127 in Sgt1 in the Structure of the Sgt1:Skp1 Complex** – The $2F_O$-F_C side chain density for many residues within the capping helix is undefined when the map is contoured at 1.48 sigma. In both chain B and chain A, Trp127 is stabilised in a hydrophobic core formed by other helices of the TPR, with chain C offering further stability via its interaction with Skp1 (see discussion). The labelling of images corresponds to the respective chain IDs, with Chain A in **Yellow**, Chain B in **Slate** and Chain C in **White**.

This density bridges the two molecules, suggesting covalent linkage of the two via a disulphide bridge. However, as this region is not in contact with the Sgt1 TPR domain, it is likely of no functional relevance and an artifact of the crystallisation process, as a result of the gradual oxidation of DTT over the course of equilibration of the crystallisation drop. This process has probably aided the formation of crystals of a complex between Sgt1 and Skp1.

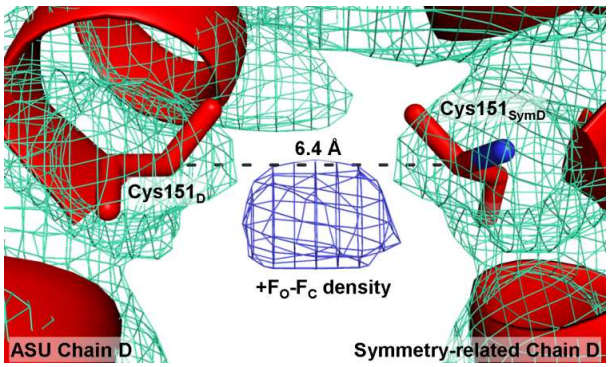


Figure 4.12: **Evidence for a Disulphide Bond Between Symmetry Related Skp1 Molecules** – During model building and refinement, a large positive F_O-F_C density developed between two cysteines present at the C-termini of Skp1. As shown here, the protein backbone is separated by 6.437 Å at the cysteine Cα. Positive F_O-F_C is shown in **Blue**, $2F_O$-F_C is shown in **Green**.

4.5 Secondary and Tertiary Structure analysis of the Sgt1:Skp1 Complex

The consensus TPR structure previously solved (Main et al., 2008) represents an interpretation of an average three-repeat TPR fold. As such, it provides a good measure for structurally classifying the Sgt1 N-terminal domain as a TPR. A direct comparison of the Sgt1 TPR with other TPR domains might also help explain why attempts at phasing previous maps by molecular replacement were unsuccessful.

Comparing the topological arrangement of the secondary structure of Sgt1 against the consensus TPR domain highlights the preservation of the helix-turn-helix repeat motif within the Sgt1 TPR (Figure 4.13). The consensus TPR and Sgt1 chain C have an alpha helical secondary structure, with 7 alpha helices comprising 77.2% of the structure of chains B and C, and 76.7% of chain A. However, differences in helix lengths and a secondary structure insertion of a 3_{10} helix (2.2% of the Sgt1 secondary structure) between TPR helices 2A and 2B, reflect how Sgt1 differs from a typical TPR in terms of secondary structure. Helices in the consensus structure are an average of 21.13 Å in length, while TPR helices of Sgt1 range in length from 17.88 Å to 32.20 Å, leading to a greater average of 22.8 Å.

A best-fit structural superposition of the core TPR domain of Sgt1 with other TPR domains lends some support to its deviation from a typical TPR fold (Figure 4.14). While the helix-turn-helix secondary structure is preserved between each of the aligned structures, the secondary structure insertion in the loops between Sgt1 TPR helix 2A and 2B and the extended conformation of helix 2A relative to other TPRs, highlights structural differentiation in Sgt1.

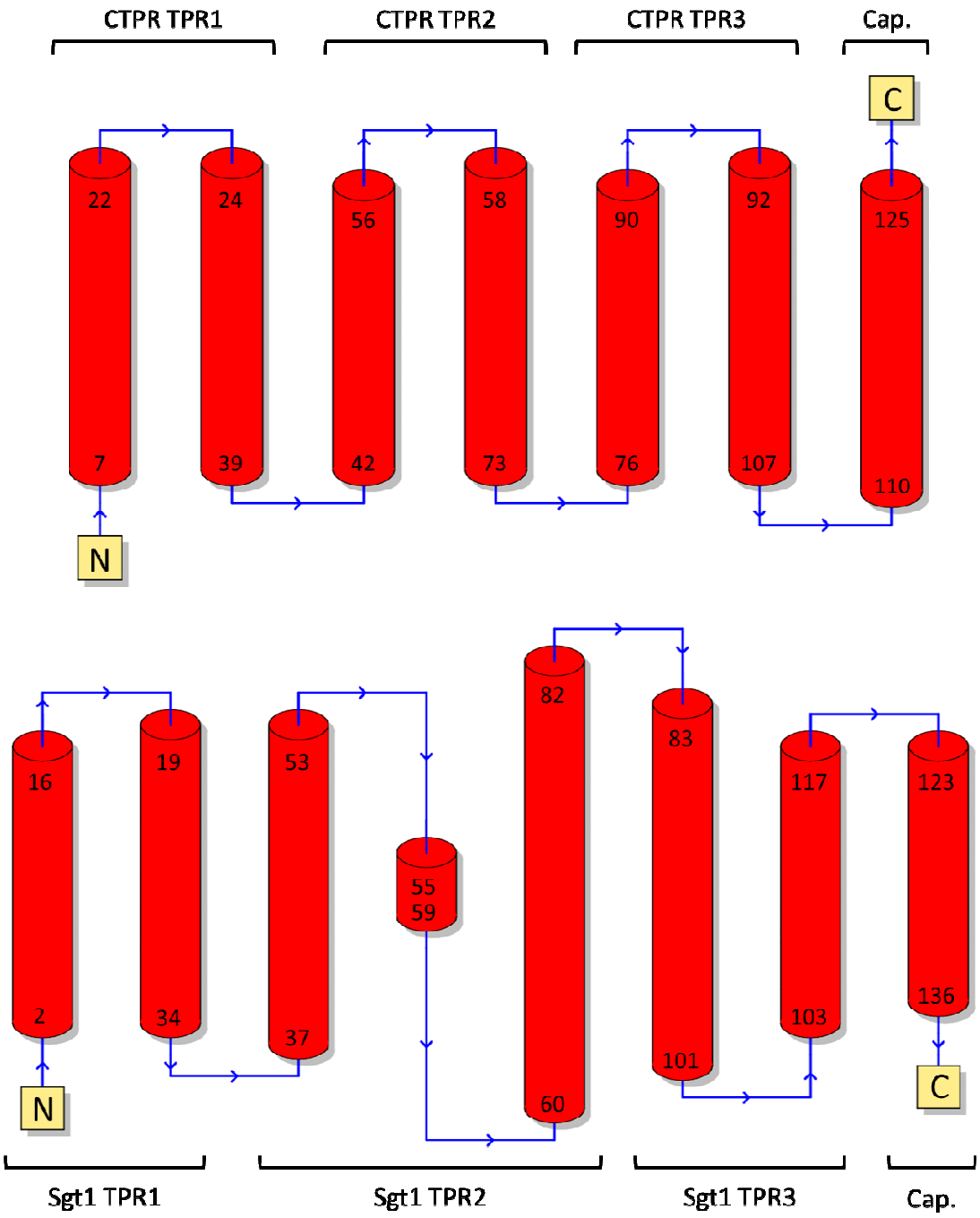
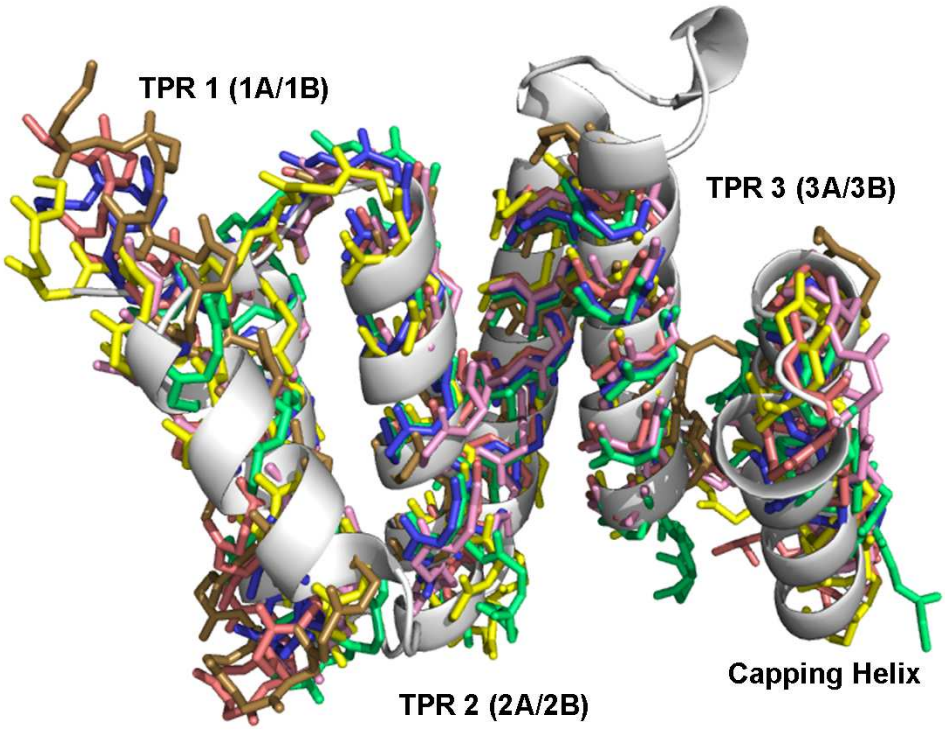


Figure 4.13: **Comparison of the Topology of the consensus three-repeat TPR and Sgt1 TPR** – An assessment of the topology of the consensus TPR (**Top**) and Sgt1 TPR (**Bottom**) highlights the symmetrical arrangement of TPRs in an average three-repeat TPR fold. In contrast, the Sgt1 TPR exhibits varied helical lengths between TPR pairs. The components of this figure were generated using PDB sum (Laskowski, 2009).

The root mean square deviation of Sgt1 C_α atoms compared to other TPR domains is reflective of the variance in intra- and inter-TPR packing conferred by the primary structure of Sgt1. While this divergence in sequence preserves the overall helix-turn-helix repeat motif (Figure 4.14), it also determines differences

in inter-TPR packing angles and the overall conformation of the three TPR repeats.



PDB ID	Protein	Colour	RMSD to Sgt1 TPR	Z-Score
-	Sgt1	White	-	-
1ELR	Hop TPR2A	Yellow	2.07	6.3
1A17	PP5	Deep Salmon	2.02	6.6
1ELW	Hop TPR1	Pink	2.13	6.6
1NA0	CTPR3	Blue	2.14	5.5
2PQN	Fis1	Sand	2.46	4.6
2XEV	YbgF	Lime Green	2.76	5.0

Figure 4.14: **Structural Superposition Sgt1 TPR Domain Against Other Known TPR Structures** – Sgt1 deviates most significantly in TPR 2, which has two elongated helices comprising this TPR repeat. This figure highlights the conservation of the overall TPR tertiary fold amongst different three-repeat TPR domains. The RMSD values were generated using the DaliLite for pairwise structural superposition (Holm *et al.*, 2000).

In a comparison of inter-TPR angles in Sgt1 and six other three-repeat TPR domains (Table 4.2), it is clear that the packing of TPR helices and the pattern of residues that is critical for the TPR fold is conserved well between the chosen proteins. Conversely, intra-TPR angles, which play a significant role in determining the contour of the convex and concave face, are significantly divergent amongst the chosen group. Angles between helices that determine intra-TPR packing range from a standard deviation of 1.03° in the bacterial TPR

protein YbgF (PDB ID: 2XEV; Krachler, Sharma and Cauldwell *et al*., 2010) to 7.48° in TPR2A of Hop (PDB ID: 1ELR; Onuoha *et al*., 2008), and 10.72° in Sgt1.

Helices	Type of Angle	Protein (PDB ID)						
		Sgt1	*CTPR3 (1NA0)*	*Hop TPR1 (1ELW)*	*Hop TPR2A (1ELR)*	*PP5 (1A17)*	*Fis1 (2PQN)*	*YbgF (2XEV)*
		Angle (180°-Absolute Angle) (°)						
1A:1B	Intra-TPR	21.12	15.43	13.95	12.99	14.23	30.02	21.07
2A:2B		14.36	17.03	20.26	13.05	17.50	26.52	21.52
3A:3B		17.97	14.38	9.94	14.71	20.65	32.32	12.44
Standard Deviation of Intra-TPR Angles		3.38	1.33	5.20	0.98	3.21	2.92	5.12
1B:2A	Inter-TPR	35.45	32.43	23.28	16.50	27.62	33.20	32.77
2B:3A		16.47	28.39	26.49	18.07	28.51	27.38	31.08
3B:Cap.		34.61	28.88	32.48	30.17	30.42	n/a	32.96
Standard Deviation of Inter-TPR Angles		10.72	2.20	4.67	7.48	1.43	4.12	1.03

Table 4.2: **Summary of Intra- and Inter-TPR Angles of Different Three-Repeat TPR Domain-containing Proteins**

Consistent packing angles between TPR repeats results in the closest points of contact being inter-helical, as is the case in CTPR3. In Sgt1, the angular twist from TPR to TPR is complemented by the inherent variance in helix lengths, and instead brings about contacts at the extremities of TPR helices, changing the residues that are accessible via the concave and convex surface of the TPR domain. The packing of the C-terminal end of helix 2B behind helix 2A, in particular, causes the protrusion of 2A further into the concave surface of the TPR, which in turn results in a distance of 13.1 Å between helix 2A and 3A at their N-terminal ends (Figure 4.15).

The summative result of this is tight packing of the respective C- and N-termini of 2B and 3A, separated at their ends by 8.8 Å. At the opposite helical pole, distances between helices 2B and 3A are increased to 10.1 Å, conferring a slightly 'splayed' structure to this half of the TPR along the convex face and a larger surface area for quaternary interactions. Indeed contacts made by helices 2B, 3A and 3B involve 44 of the 130 inter-helical contacts (34%). In this layout, the tripartite network of interactions between helices 2A, 2B and 3A form the core of the TPR, while the other helices (1A, 1B and the capping helix) form the 'walls' of the concave groove.

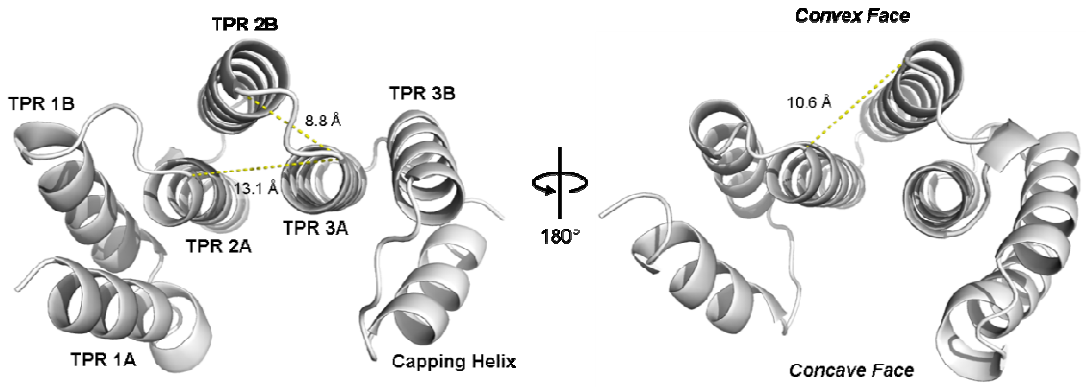


Figure 4.15: **Separation of Helices** – The packing of the helices of the individual TPR repeats leads to a slightly skewed architecture in one portion of the TPR domain. This has the effect of spreading the potential contact area on the convex face of the TPR, while the extended helix lengths permit the projection of residues into the concave face. This splayed arrangement might facilitate further interactions within the concave groove formed by the packing of TPR helices.

A feature of Sgt1 that is also present in other TPR domains is the interdigitation of the first and second helices of a TPR repeat, such that the first helices of each of the three repeats forms the contact surface for the concave face, while the second helices form the convex surface (Figure 4.16). What differentiates Sgt1 and other TPR domains from each other is how the interdigitated helices protrude from opposing faces.

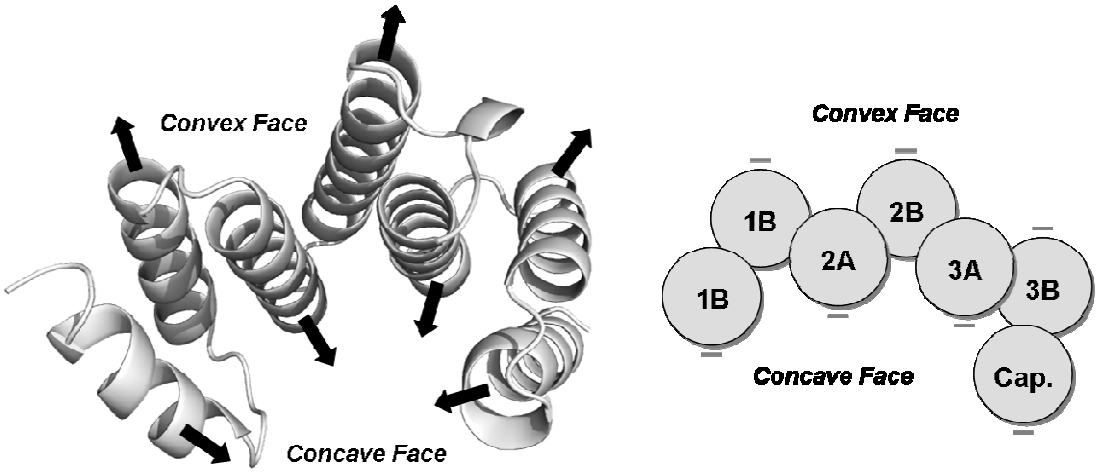


Figure 4.16: **Interdigitation of Helices and the Architecture of the Concave and Convex Faces of the TPR** – The packing of TPR helices, and packing of TPR repeats against each other, results in the protrustion of the first helix of each TPR pair into the concave face, while set of second helices forms the convex surface.

In light of the observations on the overall architecture of the Sgt1 TPR domain, it is possible that the original molecular replacement attempts failed due to the

combined differences in the packing angles between TPR repeats and the lengths of TPR helices, in addition to the unique 3_{10} helix insertion. While inter-TPR packing angles are conserved amongst three-repeat TPR proteins, it is ultimately the combination of its unique sequence and the resulting tertiary fold (including angles, packing, and the topology of the concave and convex faces) that sets Sgt1 apart from other TPR domains.

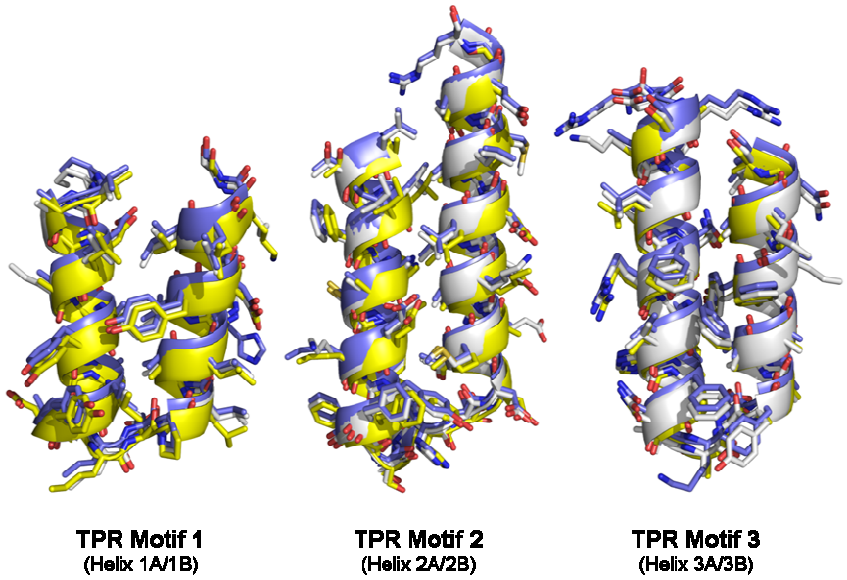


Figure 4.17: **Structural Superposition of TPR Motifs from Different Chains of Sgt1 within the Asymmetric Unit –** The TPR motifs of different chains in the structure of Sgt1 and Skp1 share similar packing and side chain orientations. This highlights the fact that despite the incompleteness of the loop regions, chains A **(Yellow)** and B **(Slate)**, have a similar overall conformation to chain C **(White)**, the most complete of the three copies of Sgt1 in the asymmetric unit. This structural superposition was generated using DaliLite (Holm *et al.*, 2000).

As a consequence of imposing NCS restraints in Buster, equivalent sets of TPR repeats between Sgt1 chains do not differ significantly in their pitch, orientation and side chain position. A structural superposition of the TPRs shows that the individual repeats overlay well, and that side chain orientations are largely matched between the three chains (Figure 4.17). Because of the relatively low resolution, it is difficult to conclude on differences in side chain orientations and preferred side chain rotamers. A higher resolution data set is therefore required before it would be possible to draw conclusions on changes to residue orientations caused by ligand binding (for example, Skp1 binding).

In both the consensus TPR and Sgt1, the capping helix is of approximately equivalent length to the very C-terminal helix (helix 3B) of the core TPR fold (19.59 Å and 17.88 Å in Sgt1), which likely facilitates its function in stabilizing the TPR domain and enhancing its solubility *in vitro*. Helices 3A and 3B, for example, have a number of phenylalanines that would be in unfavourable positions in the absence of the capping helix. Leu131 and Trp127 provide a more favourable environment when within their proximity (Figure 4.18). In addition, this hydrophobic core accommodates Tyr119, which projects from the 3B-capping helix loop into the interface formed by 3A and the capping helix. It is probably for this reason that the previously discussed Sgt1 construct, truncated at residue 121, is insoluble.

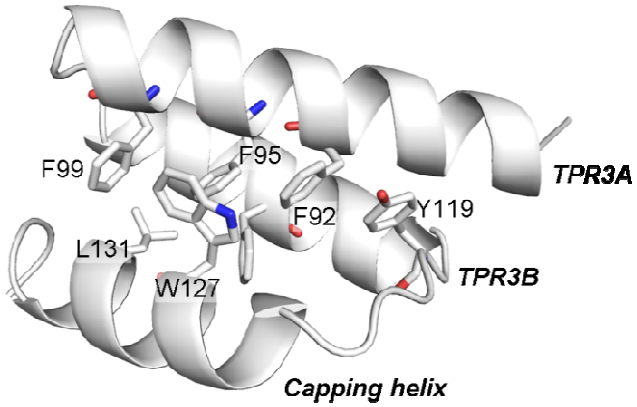


Figure 4.18: **Packing of the Capping Helix against Helix 3B of the Sgt1 TPR Domain –** Packing of the capping helix against the last TPR repeat stabilises a number of hydrophobic residues that, in its absence, would otherwise be in unfavourable environments. Trp127 and Leu131, in particular, facilitate solvent exclusion when packed against Phe99 and Phe111 of the adjacent helices. Tyr119 also contributes to this interface, projecting from the 3B-capping helix loop in to the hydrophobic core.

4.6 Structural Features of the Sgt1:Skp1 Interface

The structure of the Sgt1:Skp1 complex allows a molecular understanding of two aspects of Sgt1's function *in vivo*: its dimerisation and its association with Skp1. The following discussion will review the interactions seen in the structure at the Sgt1:Skp1 interface.

The Sgt1 molecule interacting with Skp1, chain C, contacts the BTB/POZ domain via the concave groove of the TPR (Figure 4.19, A). Skp1 itself binds to one half of the TPR, projecting outward at an angle of 63.2° from helix 3A.

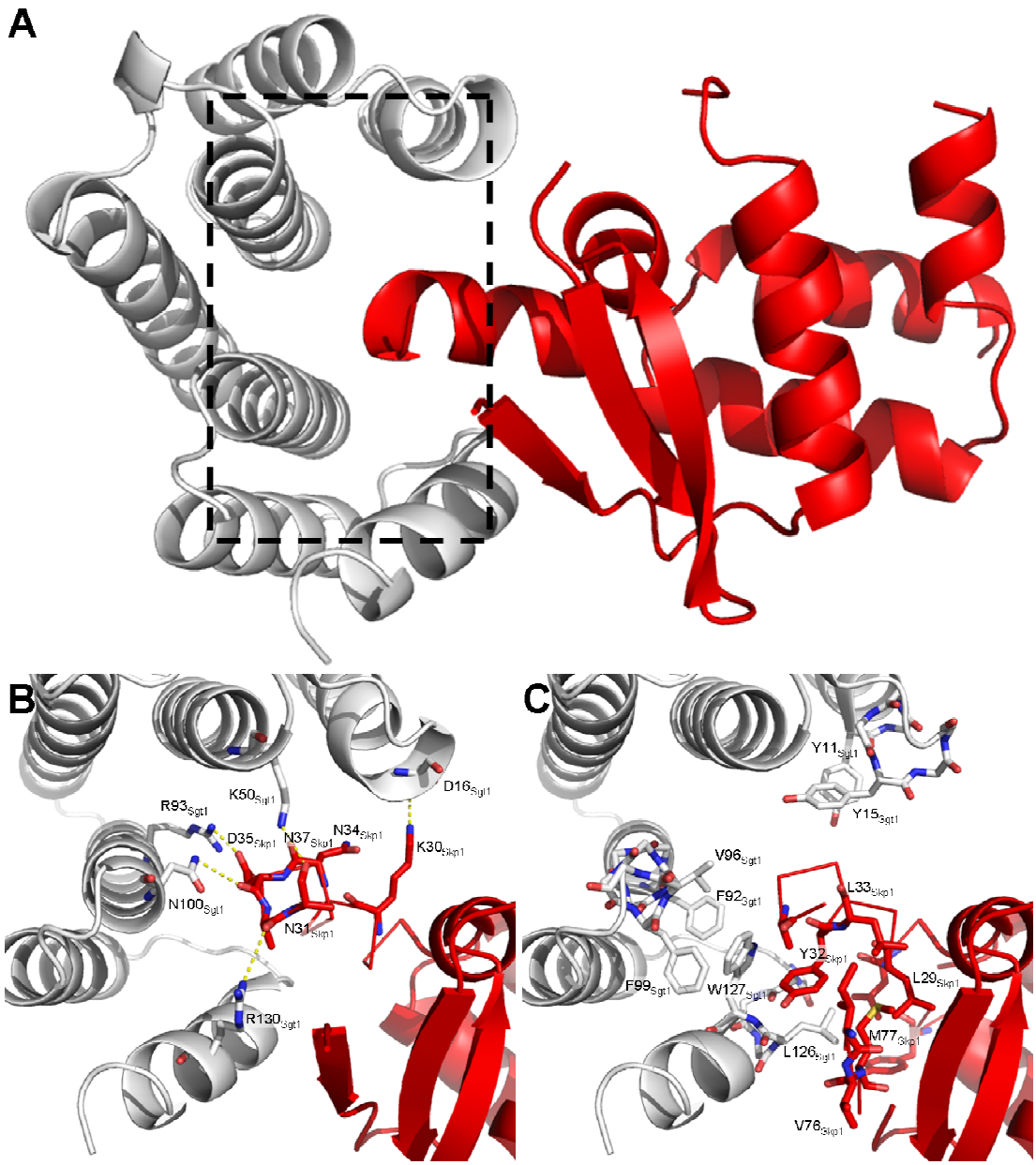


Figure 4.19: **Summary of the Sgt1:Skp1 Interface** – **(A)** The Sgt1:Skp1 interface in cartoon representation. Skp1 binds in the concave groove of the TPR domain. **(B)** All non-hydrophobic residues that take part in the interface shown in stick representation. Hydrogen bonds and salt bridges are shown for interacting atoms that are no more than 3.9 Å apart. **(C)** Hydrophobic residues that are involved in the interaction in stick representation.

The Sgt1:Skp1 interface seen in the structure buries approximately 600.2 Å^2 of surface area on Sgt1 and features contributions from four Sgt1 helices, which bury 626.4 Å^2 of surface area on Skp1. At the center of these interactions is a salt bridge between Arg93 residue on Sgt1, present on helix 3A, and Asp35 on Skp1, which is reinforced by further hydrogen bonds contributed from Lys50, Asn100 and Arg130, and a peripheral salt bridge between Asp16 on Sgt1 and Lys30 on Skp1 (Figure 4.19, B). The interface is anchored by a small region of hydrophobic

packing on the side of the TPR capping helix, with critical contributions from Leu126 (about 121.40 Å^2) and Trp127 (about 55.78 Å^2) (Figure 4.20).

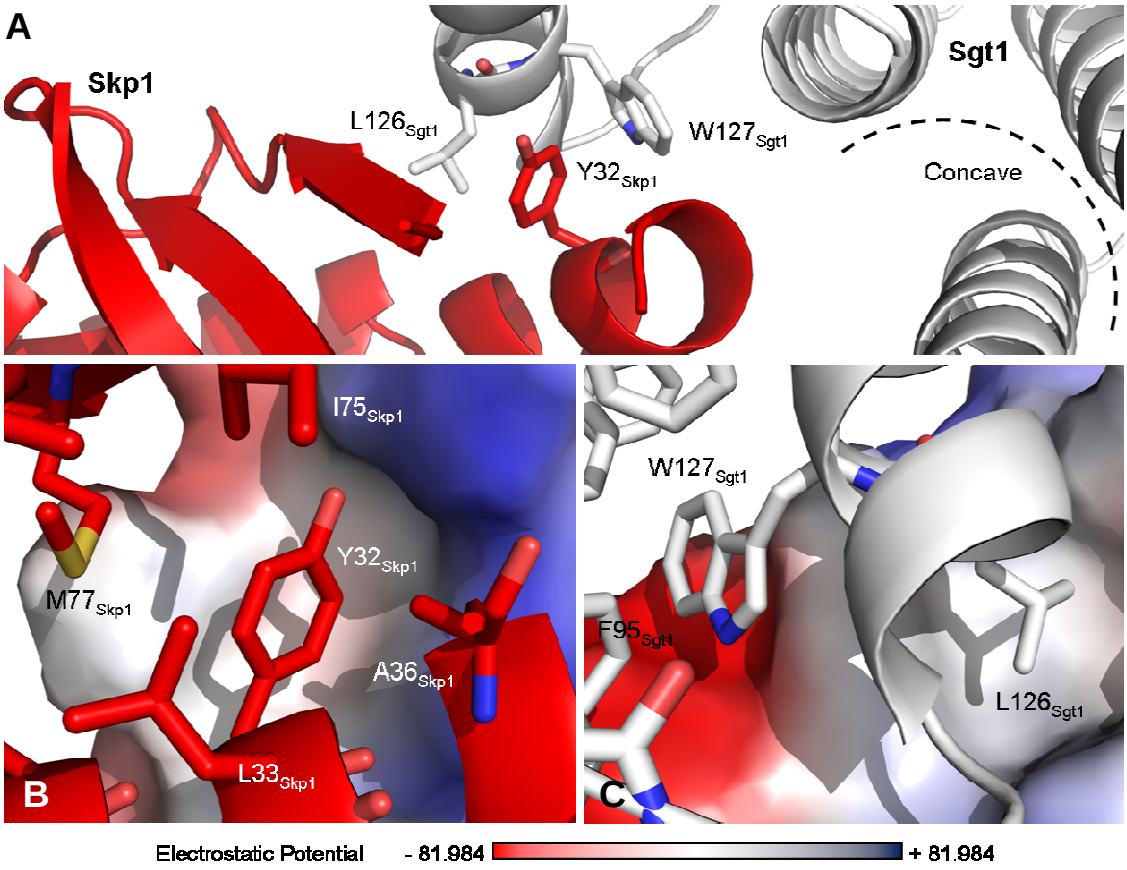


Figure 4.20: **Close-up of the Hydrophobic Core of the Sgt1:Skp1 Interface – (A)** The hydrophobic anchor of the Sgt1:Skp1 interaction occurs on the side of the Sgt1 capping helix, with Sgt1 contributing two residues (Leu126 and Trp127) at this site of interaction **(B)** The tyrosine contributed by Skp1 (**Cartoon**, **Red**) to this interface is buried within a groove formed by residues within the capping helix. **(C)** Leu126 buries in a pocket of no net charge in Skp1, while Trp127 provides a surface for edge-on packing of the side chain of Tyr32 of Skp1.

A survey of Sgt1 sequence conservation between yeast and higher shows poor overall conservation of *S. cerevisiae* Sgt1 to Sgt1 in higher eukaryotes, but significant conservation amongst yeast homologues (Figure 4.21, A). The Sgt1:Skp1 interface (Figure 4.21, sticks) is functionally more conserved in yeast compared to higher eukaryotes. At the centre of the concave face lies Arg93, which is highly conserved in all eukaryotes, and is flanked on either side by residues that are conserved only in yeasts that form the 'walls' of the concave TPR groove (including Tyr11, Tyr15, Asp121 and Thr123). This residue forms a central hydrogen bond with Asp35 on Skp1 (Figure 4.19, B).

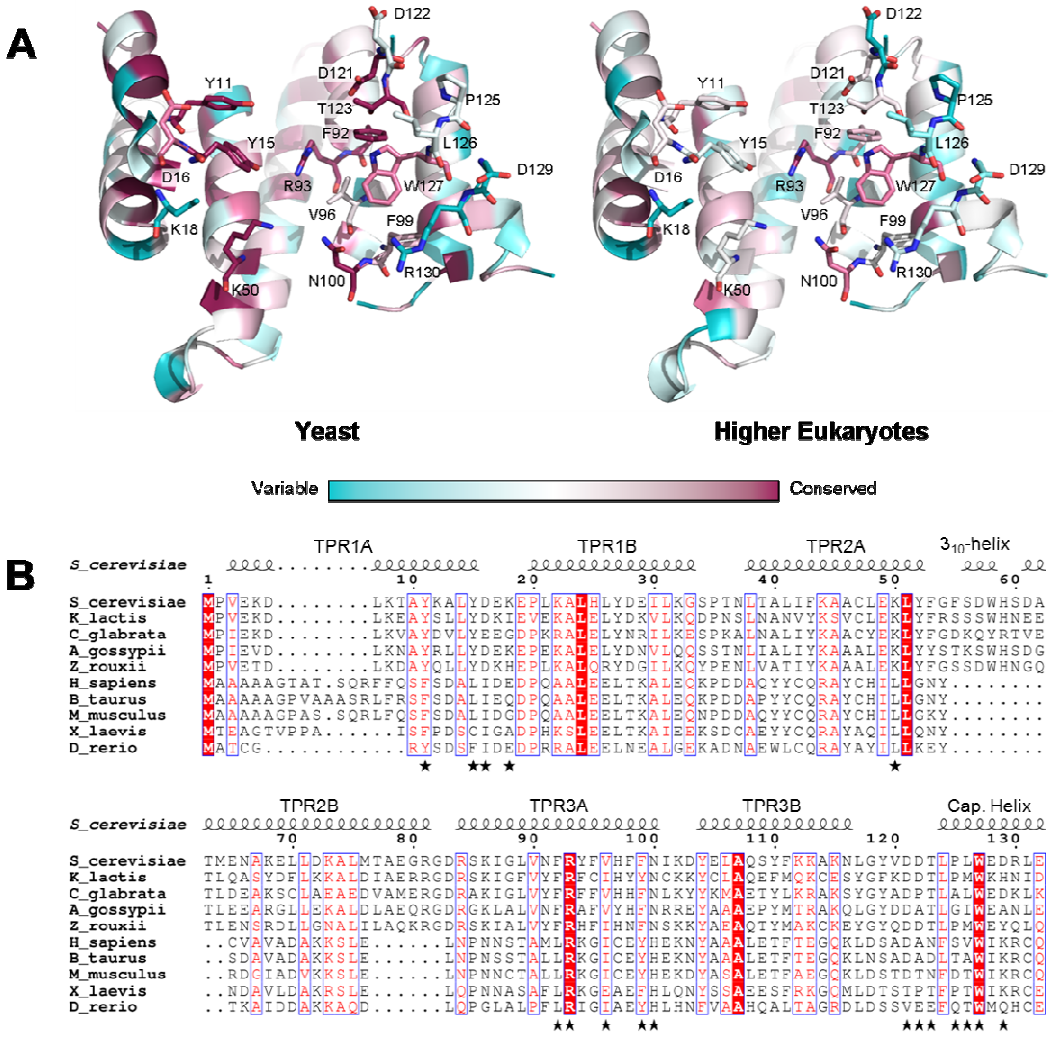


Figure 4.21: **Conservation of Residues on Sgt1 Involved at the Sgt1:Skp1 Interface – (A)** Residues involved at the Sgt1:Skp1 interface are well conserved in terms of identity in yeast, but less conserved in higher eukaryotes. The conservation is coloured by the degree of residue identity, as generated by the ConSurf server (Ashkenazy *et* al., 2010), which estimates the evolutionary conservation of residues based on the phylogenetic relationship between related sequences. **(B)** Structured-based sequence alignment of Sgt1 orthologues generated by T-COFFEE EXPRESSO (Notredame *et al.*, 2000), emphasising the conservation of residues involved in the Sgt1:Skp1 interaction (marked by a ★). This figure was generated using ESPript (Gouet *et al.*, 1999).

In the capping helix, Trp127 is conserved and lies at the centre of a hydrophobic region that also includes the yeast-specific residue Phe92. Residue conservation in yeast is distributed evenly between the N- and C-terminal halves of the TPR, and more hydrophobic residues occur past the 3_{10} helix insertion than before it.

A closer look at the structure-based sequence alignment shows that the residues involved at the Sgt1:Skp1 interface in yeast are substituted for appropriate

residues in higher eukaryotes (Figure 4.21, B), which infers that the interface for Skp1 binding is likely to be the same in these organisms.

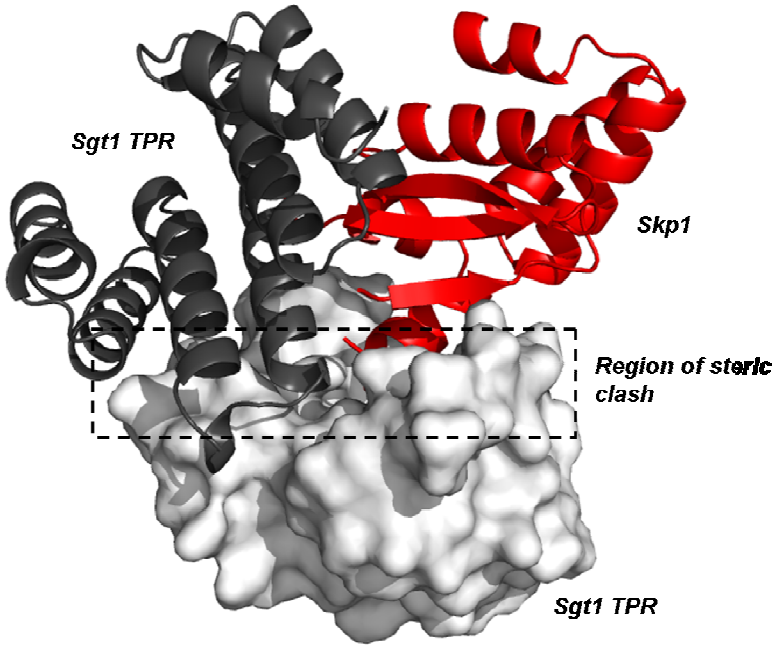


Figure 4.22: **Shared Interfacial Residues between the Sgt1 homodimer and the Sgt1:Skp1 Heterodimer –** Sgt1 (**Black**) and Skp1 (**Red**) overlap in their binding sites (**Dashed Box**) within the concave groove of any given Sgt1 protomer (**Surface**, **White**). This steric clash prevents binding of Skp1 and homodimerisation via the concave binding site of a 'free' Sgt1 TPR protomer; it is possible that some competition occurs for these sites in solution. The reverse convex face of the host Sgt1 protomer remains free to interact with another TPR protomer, thereby maintaining a 2:1 stoichiometry. The structural superposition utilised to generate this image was performed in DaliLite (Holm *et al.*, 2000).

Chain D makes an additional contact with the convex face of a symmetry-related chain C molecule, though this interface is significantly smaller, with an area of 243.5 Å^2 on chain D buried by this interaction. This interface comprises fewer salt bridges, with Glu122 on Skp1 interfacing exclusively with Lys68 and Lys113 from Sgt1. A structural superposition with a Skp1 structure bound to an F box-protein (data not shown) shows that the position of the Sgt1 protomer in this interface would interfere with the binding of an F box-containing protein, such as the CBF3 kinetochore component Ctf13. Combined with the lack of sequence conservation at this interface, it is therefore unlikely that it bears any biological significance.

Within the asymmetric unit, the Skp1 binding site on the protomer contributing its concave face to a dimer of Sgt1 is occupied (e.g. chain A, in the C:A interaction) and therefore not available for Skp1 binding. Superimposing the Sgt1:Skp1 interaction onto an Sgt1:Sgt1 pair confirms the overlap of interaction sites (Figure 4.22). This explains why the interaction with Skp1 that is visible in the model is fundamentally 1:1, since an Sgt1 protomer and Skp1 cannot simultaneously bind the Sgt1 TPR domain via the same site. In this mode of dimerisation, Sgt1 maintains one available binding site, while its binding partner is unable to form further interactions via the concave face. The implications of this will be discussed further in Section 5.0.

In summary, the interface between Sgt1 and Skp1 is significant, albeit relatively small, with half of the interface hydrophobic and half hydrophilic. The charge-charge interactions predominantly occur via the N-terminal half of the TPR domain before the 3_{10} helix insertion, and the hydrophobic core occurs on the side of the capping helix. In comparison to other domain:domain interactions, such as the interaction of the Skp1 BTB/POZ domain with Cdc4 (2762.2 Å^2) (PDB ID: 3MKS; Orlicky *et al*., 2010) or the Sgt1 CS domain complexed with the N-terminus of Hsp90 (1669.4 Å^2) (PDB ID: 2XCM; Zhang *et al*., 2010), the Sgt1:Skp1 buries only a relatively small amount of total surface area between the interacting molecules (1226.6 Å^2). The interaction is similar in size to the Skp1:Cul1 interaction (1373.9 Å^2) (PDB ID: 1LDK), which might have some functional relevance, since Sgt1 and Cul1 share a binding site on Skp1 and compete for binding *in vitro* (Zhang *et al*., 2008).

4.7 Structural Features of the Sgt1:Sgt1 Self-Association Interface

The structure of Sgt1 highlights two possible modes of self-association: a stacking interaction that doubles the accessible surface area of the concave TPR face and a novel mode of TPR self-association involving the loop insertion of the TPR.

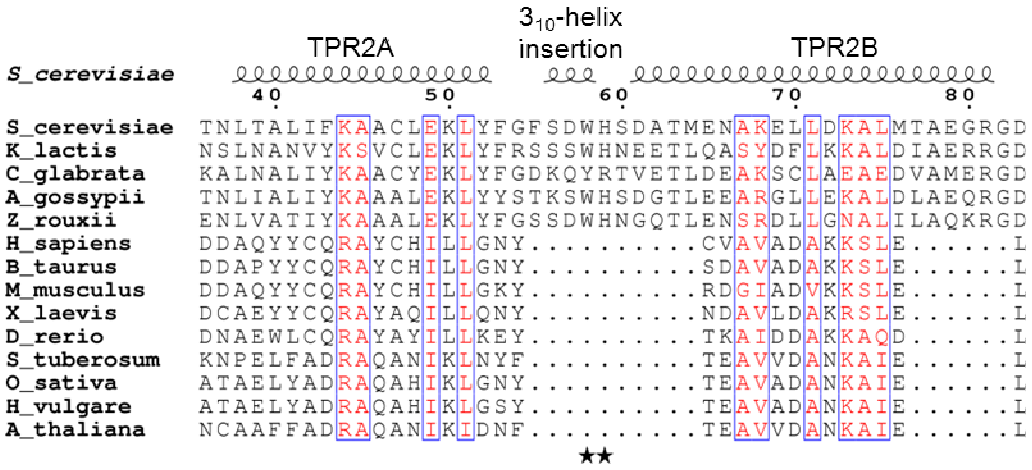


Figure 4.23: **Structure-based Sequence Alignment of Sgt1 Homologues** – A structure-based sequence alignment of Sgt1 homologues using T-COFFEE EXPRESSO (Notredame *et al.*, 2000) shows the divergence in sequence between yeast and higher eukaryotes that might confer a species specific function to the TPR domain of Sgt1. In particular, the yeast specific insertion in the loop between TPR 2A and 2B is visibly involved in dimerisation in the *S. cerevisiae* homologue, but is not present in higher eukaryotic species. The residues involved in the key stacking interaction at this interface are marked by a ★. Helix 2B also appears more elongated in yeast than in other eukaryotes. While it is known that plant Sgt1 forms dimers and the human homologue does not (Nyarko *et al.*, 2007), the oligomerisation behaviour of other homologues in this sequence alignment is unknown. The alignment image generated using ESPript (Gouet *et al.*, 1999).

A sequence alignment of Sgt1 homologues, with an emphasis on yeast orthologues, underlines the sequence divergence of Sgt1 across species and points to yeast-specific sequences that correspond to the aforementioned secondary structure insertion between TPR helices 2A and 2B (Figure 4.23). This insertion comprises a number of hydrophobic residues that are involved at the Sgt1:Sgt1 interface, pivotal to which is a pi-stacking interaction between Trp 58 from one monomer (e.g. chain C, white; Figure 4.24, D) and His 59 from the protomer packing against its convex face (e.g. chain A, yellow; Figure 4.24, D). Peripheral to this are a number of electrostatic interactions from adjacent helices that stabilise this dimer, including a salt bridge between Asp57 and Asp61 of one TPR and His59 and Arg130 of its partner (Figure 4.24, C). The interaction between Asp57 and His59 appears to orientate the histidine to facilitate the stacking interaction between it and Trp58 (Figure 4.24, E).

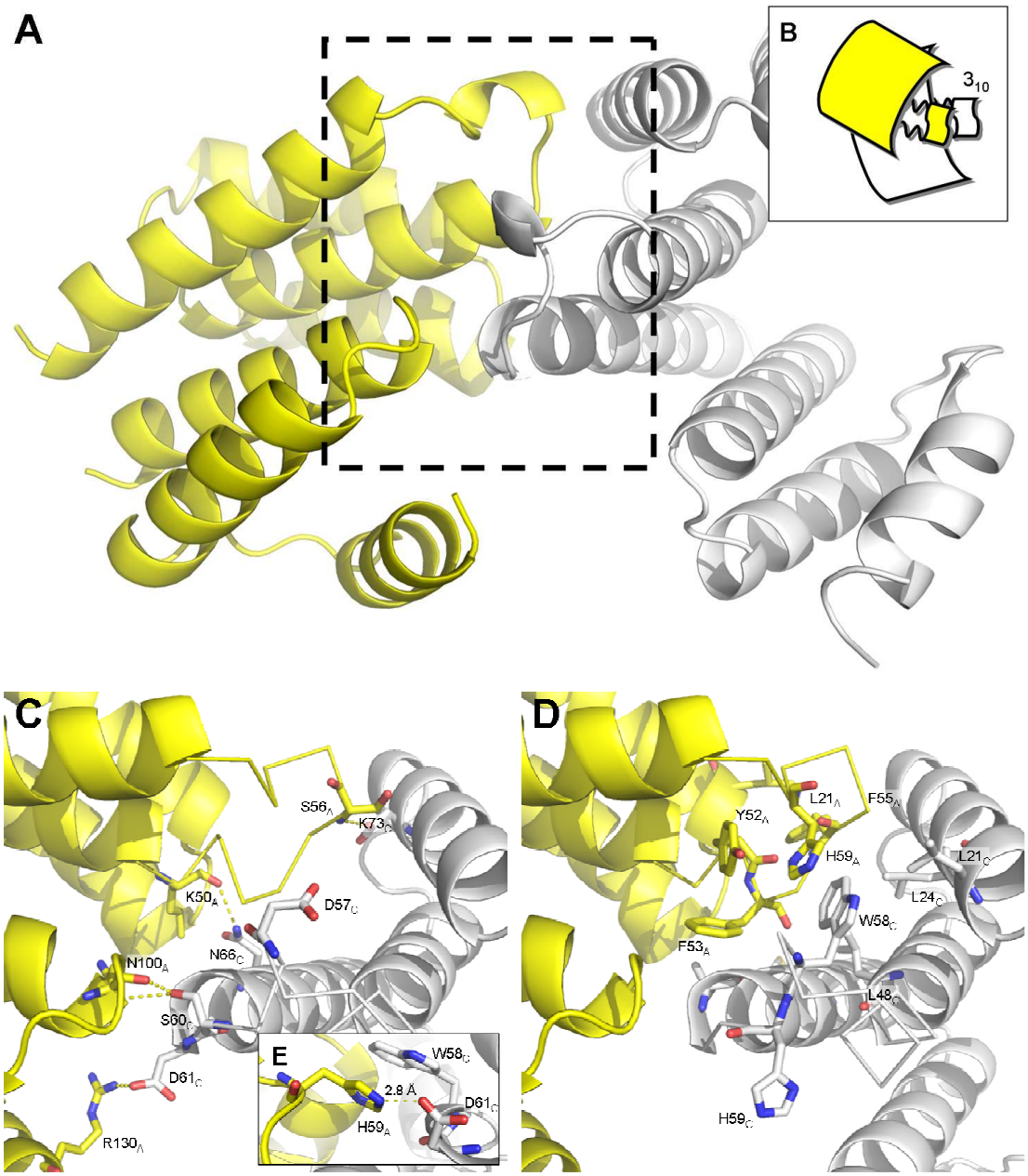


Figure 4.24: **Summary of the Sgt1:Sgt1 Interface Seen in the Asymmetric Unit – (A)** Cartoon representation of the Sgt1 dimer interface seen in the crystal structure of the Sgt1:Skp1 complex. The interface within the **Dashed Box** region. **(B)** Stylised depiction of the interface highlighting the interaction at the yeast-specific 3_{10} helix insertion. **(C)** Shown here is the complete set of non-hydrophobic residues involved at the Sgt1:Sgt1 self-association interface formed between Sgt1 chains A and C, and B and A, with the interacting residues shown in stick representation. Hydrogen bonds and salt bridges are shown for interacting atoms that are no more than 3.9 Å apart**. (D)** All hydrophobic residues that are involved at the interface shown as sticks. **(E)** Orientation of His59 by Asp57 for stacking with Trp58.

As a result of this interface, Trp58 and His59 point in opposite directions in the intra-TPR2 loop (Figure 4.25), exposing Trp58 on A for an identical interaction with the third Sgt1 protomer in the asymmetric unit. The interaction of C:A is therefore mimicked on the opposite face of A, via an interaction with chain B. In

turn, both the core and peripheral interactions of the two dimer interfaces visible within the asymmetric unit are largely identical, burying a total of 1212.9 Å^2 of surface area in the C:A interface (596.2 Å^2 in C and 616.7 Å^{2}in A), and 1257.1 Å^2 the A:B interface (613.3 Å^2 in A and 643.8 Å^2 in B).

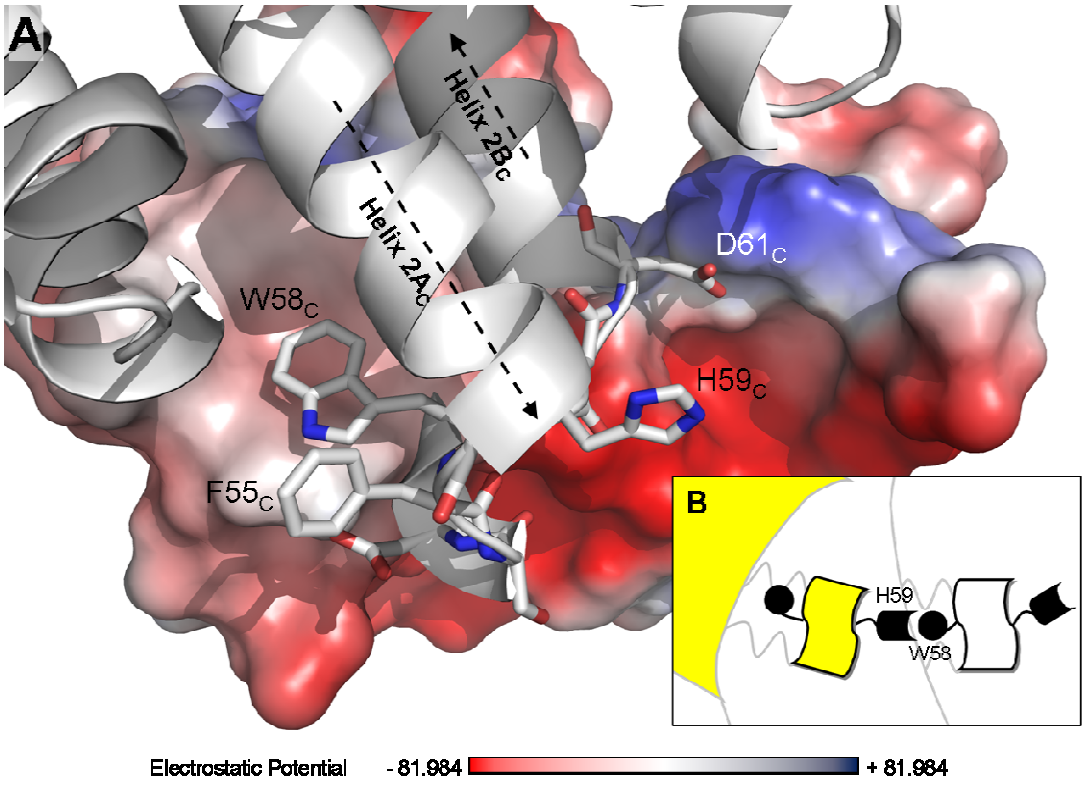


Figure 4.25: **Close-up of the Sgt1:Sgt1 Dimerisation Interface –** One of the dimer interfaces within the crystal structure involves a stacking interaction between His59 from one TPR protomer and Trp58 from its adjacent partner. Trp58 from the cartoon-represented protomer is shown to pack in a groove formed by the conserved loop region. His 59 on the same chain follows the helical axis and is orientated in the opposite direction to Trp58.

While this mode of dimerisation theoretically opens the free-histidine protomer to an equivalent interaction in chain C, the majority of the interfacial residues are buried by the interaction of chain C with Skp1.

Previous analyses of Sgt1 1-150 by SV-AUC highlight the tendency of this construct to form trimers in solution. The arrangement of TPR domains within the asymmetric unit sheds light on a possible mode of trimerisation, with TPR protomers packing via the conserved yeast loop region to form a trimer. However, although the donating tryptophan (Trp58) on the last protomer within the trimer is free to facilitate a tetramer (and, from there, higher order

oligomers), and the crystallisation concentrations are significantly higher than those likely to be found of Sgt1 *in vivo*, these types of oligomers are not evident in the crystal structure.

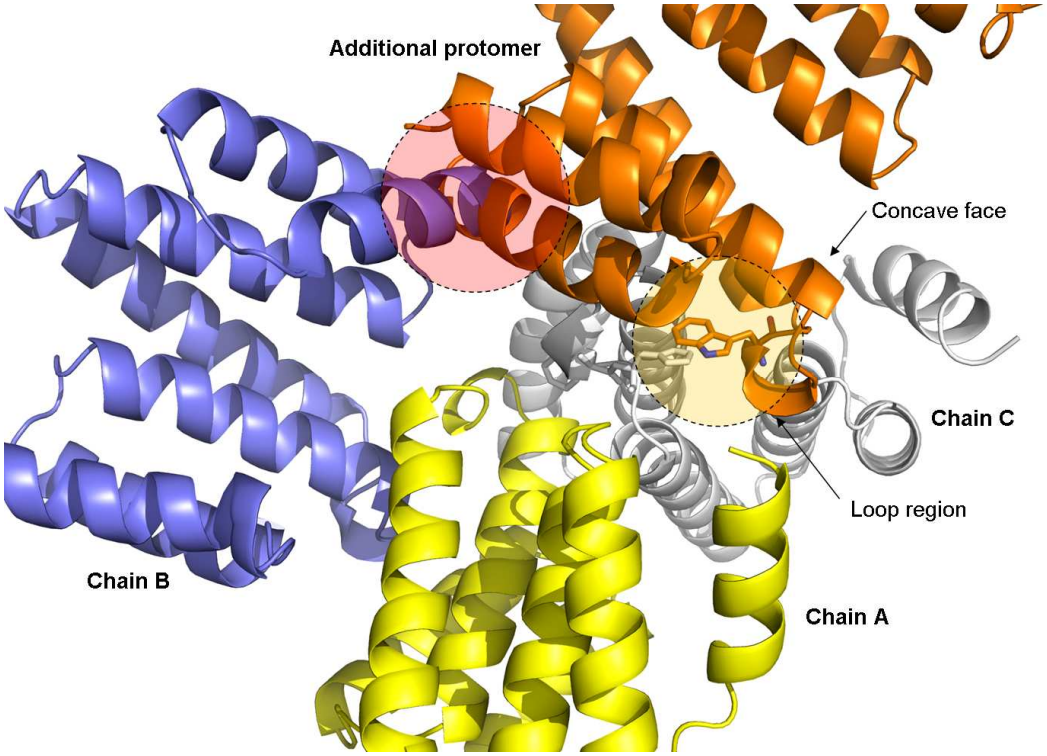


Figure 4.26: **Structural Factors Limiting Further Oligomerisation –** If a fourth TPR domain forms identical contacts via the conserved region (**Yellow Highlight**) and is placed within the existing oligomer arrangement, it clashes with other protomers in the oligomer (**Red Highlight**). In this instance, helix 2A of the **Orange** TPR protomer (placed from a structural superposition of the TPR dimer) sterically clashes with the capping helix of the **Blue** coloured protomer (chain B in the asymmetric unit). The structural superposition of TPR protomers was performed in DaliLite (Holm *et al.*, 2000).

Adding an additional TPR protomer to the existing arrangement, results in a physical clash between helix 2A of the added protomer and the trimer seen in the asymmetric unit (Figure 4.26). Furthermore, additional protomers would block the concave groove of the Sgt1 monomer that binds Skp1. This scenario would be exacerbated in the full length protein, where the presence of the CS and SGS domains of Sgt1 would likely prevent this kind of oligomerisation.

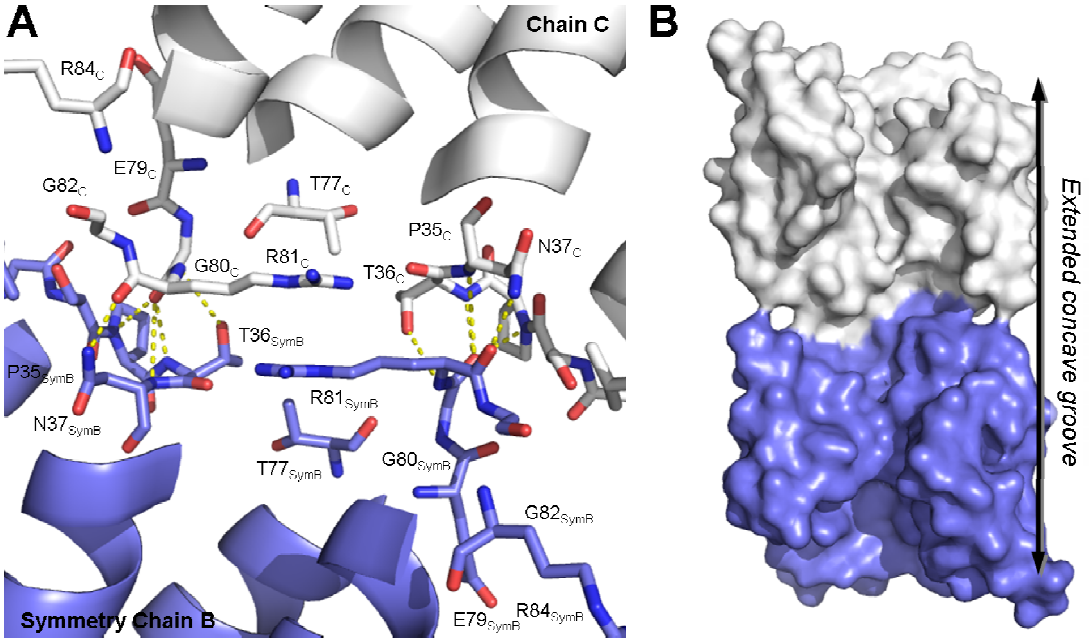


Figure 4.27: **Secondary Interface Between Two Sgt1 Protomers – (A)** A second dimerisation interface exists across a two-fold symmetry axis, involving residues in the loop regions of the TPR domain. Inter-molecular hydrogen bonds are shown stabilizing the interaction; equivalent bond pairs are contributed from both chains. **(B)** The interfacing protomers generate an extended concave groove, which effectively doubles the binding surface available in the dimer for binding partners such as Skp1.

A second plausible interface for self-association occurs via the loop-regions of the TPR domain through contacts made by chain C of the asymmetric unit and a symmetry-related chain B. The two chains are related by a two-fold symmetry axis (Figure 4.27, A). Contributing side chains are mirrored across this interface, with backbone hydrogen bonds forming between Pro35, Thr36, Asn37 and Gly80 on both interacting chains, as well as hydrogen bonds contributed by side chains of Thr36, Asn37 and Arg81. The latter contributes approximately a third (108.64 Å^2) of the buried surface of chain C (380.5 Å^2) and chain B (382.9 Å^2). It is interesting to note that this interface would effectively double the binding surface of the concave groove to facilitate further tertiary interactions; this conformation, in effect, would allow a second Skp1 molecule to bind this TPR protomer as the binding region is on the opposite pole of the dimer (Figure 4.27, B).

4.8 VALIDATION OF THE SGT1:SKP1 AND SGT1:SGT1 INTERFACES

A structural analysis of Sgt1 and an Sgt1:Skp1 complex has provided valuable insight into the interactions that might bear biological significance. Having identified the key interfaces within the complex, it was necessary to validate these in solution and in context of the full length proteins.

Molecule	Residue	Mutation	Interface	Interfacing residue(s)	Expressed?	Purified?
Sgt1	Leu 48	Arg	Sgt1	Gly 54, Phe 55	Yes	No
	Trp 58	Ala		Asp 57, His 59	No	n/a
	His 59	Ala		Trp 58	Yes	Yes
	Asp 61	Arg		Arg 130	Yes	Yes
	Arg 93	Ala	Skp1	Asp 35	Yes	Yes
		Asp			No	No
	Leu 126	Ala		Tyr 32	Yes	Yes
	Trp 127	Ala			Yes	Yes
Skp1	Tyr 32	Ala	Sgt1	Leu 126, Trp 127, Arg 130	No	n/a
	Asp 35	Arg		Arg 93, Asn 100	No	n/a

Table 4.3: **Summary of Mutagenesis Targets on Sgt1 and Skp1 –** Each mutation was selected on the basis of its importance to the respective interfaces. Following PCR amplification, the successfully amplified targets were then sub-cloned into pET28a for large scale overexpression in T7 Express LysY/I^Q cells.

Despite the fact that the interfaces are relatively small, both the self-association site of Sgt1 and the Sgt1:Skp1 interface have a small hydrophobic anchor that appears to be critical for mediating protein:protein interactions. Table 4.3 provides a summary of the Sgt1 and Skp1 residues targeted for mutagenesis and information regarding the expression and purification of the mutants. The crystal structure shows that the N-terminus of Sgt1 is remote from both the Skp1 binding and self-association interface. Sgt1 mutants were therefore expressed with an N-terminal His_6-tag for a pulldown assay with wild type (WT) Skp1. A preliminary expression trial showed that each mutant could only be successfully expressed in the T7 Express LysY/I^Q cell strain (data not shown), with the exception of the W58A mutant, which could not be expressed. With the exception of the L48R Sgt1 mutant and Skp1 mutants, sufficient quantities of each mutant could be expressed and purified.

A pulldown assay was conducted to probe the interfaces seen in the crystal structure, with the expectation that mutations at the hydrophobic interface involving Tyr32 of Skp1 and Leu126 and Trp127 of Sgt1 would abolish complex formation. A 1.5 molar excess of untagged Skp1 was added to His_6-tagged Sgt1 to enhance complex formation.

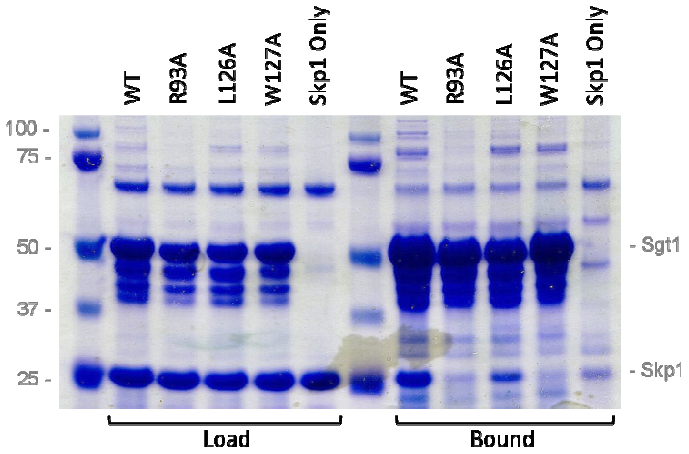


Figure 4.28: **Pulldown Assay of Sgt1 Mutants Probing the Interface with Skp1 –** The interaction between Sgt1 and Skp1 was assessed relative to the control. Wild type (WT) Sgt1 shows a significant interaction with Skp1. The L126A mutant retains some binding capacity for Sgt1, but is diminished relative to the wild type. Mutations to Arg 93 and Trp 127 completely abolish Skp1 binding. For both the wild type and mutant Sgt1 proteins, degradation products are shown by the SDS-PAGE analysis.

As expected, mutations of either of the two key hydrophobic residues at the Sgt1:Skp1 interface affects complex formation. While Leu126 on Sgt1 buries more surface area within the structure, the large aromatic side chain of Trp127 appears to have a greater influence at the interface, as shown by the abolished binding to Skp1 (Figure 4.28, load versus bound W127A). In order to probe the importance of Leu126, it might be necessary to introduce a charged residue to offset the hydrophobic interactions. Arginine 93 – the highly conserved residue at the center of the concave face of the TPR – plays a crucial role in binding Skp1, as mutating this residue to an alanine prevented Skp1 binding. The interface seen in the crystal structure therefore reflects the true interaction of the full length proteins in solution, and the functional complex *in vivo*.

The importance of dimerisation in kinetochore assembly is underscored by experiments that reconstituted self-association of monomerised Sgt1 through fusion to a non-native dimerisation domain (Bansal *et al.* 2009). This chimera was capable of rescuing yeast grown at permissive temperatures via successful kinetochore formation, while the monomerised L31P mutant was recalcitrant to

cell viability at permissive and non-permissive temperatures. Targeting residues at the interface seen within the crystal structure sheds light on a possible mode of dimerisation in solution.

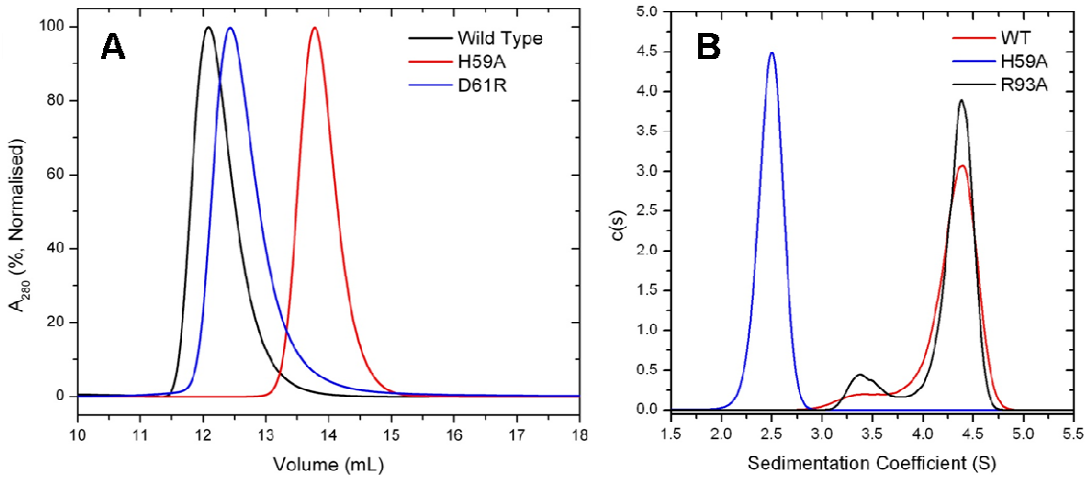


Figure 4.29: **Biophysical Characterisation of Sgt1 Dimerisation Mutants – (A)** Size exclusion chromatography profiles for wild type Sgt1 (**Black**), and H59A (**Red**) and D61R (**Blue**) mutants. The wild type protein elutes at approximately 12 mL, with the D61R eluting at approximately 12.5 mL, suggesting that it is affected in its hydrodynamic properties and possibly in its capacity to self-associate. The H59A mutant is significantly affected, with an elution volume of approximately 14 mL. **(B)** Sedimentation velocity analytical ultracentrifugation of Sgt1 mutants H59A and R93A, with R93A as an internal control. The wild type protein has a sedimentation coefficient of approximately 4.4 S, which is mirrored by the R93A mutant, suggesting that it behaves similarly to the wild type in solution. The molecular weight estimate corresponds to a dimer for the main peak. The H59A mutant is shown to be significantly affected in its hydrodynamic properties, with a molecular weight estimate corresponding to a monomer. Histidine 59 is therefore involved at the dimerisation interface of Sgt1.

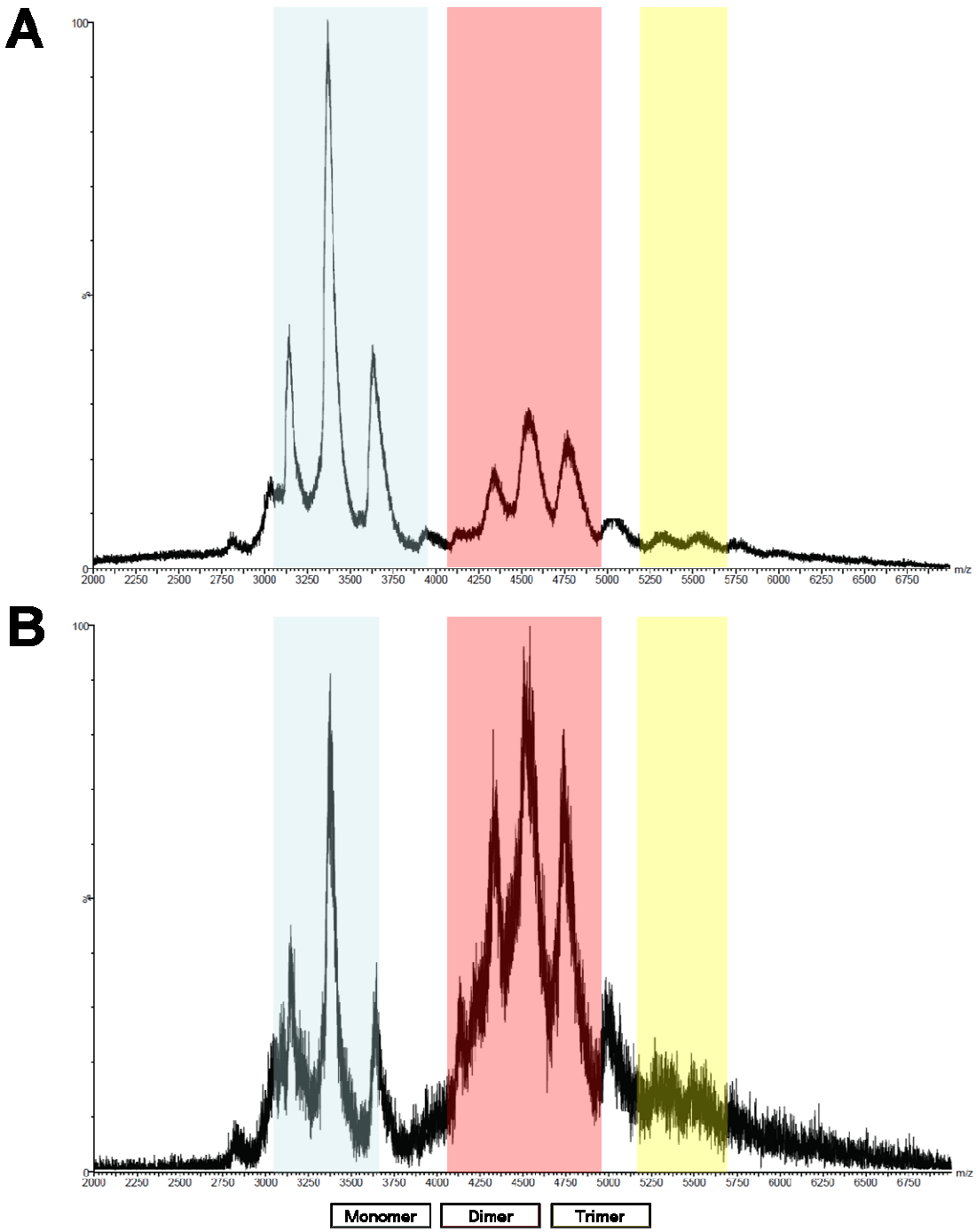
Analytical size exclusion chromatography (SEC) was used to determine the behaviour of the H59A and D61R Sgt1 mutants in solution relative to the wild type. Both mutations target residues at the dimerisation interface seen in the crystal structure of the Sgt1:Skp1 complex, which involves residues in the conserved 3_{10} helix insertion. Histidine 59 is central to this interface, while aspartate 61 lies peripherally to the main site of interaction. If the interface in the crystal structure is real, then both mutations should abolish self-association and thereby confer a different hydrodynamic behaviour to both recombinant proteins.

Both the D61R mutant and the wild type show a single peak elution profile in SEC, with a slight tail to the peak. The wild type protein has an elution volume of

approximately 12 mL (Figure 4.29 A, black); as the protein has an elongated shape, as suggested by SV-AUC, it was not possible to accurately determine the molecular weight of this species using this technique. The D61R mutant is mildly affected in its hydrodynamic properties, reflecting the peripheral nature of the interaction mediated by Asp61 in the wild type protein. The H59A mutation, however, significantly increases the elution volume of the analyte. Furthermore, the peak follows a Gaussian trend more closely than its counterparts in this experiment, suggesting that the sample is homogeneous. The difference in elution volumes confirms that specific targeting of this interface by mutagenesis has influenced the solution properties of Sgt1. The stacking interaction between Trp58 and His59 therefore participates in the dimerisation of full length Sgt1 in solution.

To further confirm the disruption of this interface, an SV-AUC experiment was conducted on the wild type and H59A mutant (Figure 4.29, B). With a sedimentation coefficient of approximately 2.5 S, the distribution profile of the H59A mutant shows a single species with a predicted molecular weight of a monomer. The wild type protein has a similar profile as in previous experiments, including the secondary peak that might correspond to an inactive monomeric species, albeit with a slightly higher S-value due to the addition of the N-terminal His_6-tag. As a control, a mutation at the Sgt1:Skp1 interface (R93A) was also sedimented, and shows a similar behaviour to the wild type. This experiment confirms the relevance of the crystal interface in solution and the mode of self-association *in vivo*.

The biophysical experiments on Sgt1 mutants infer that an Sgt1 monomer is stable in solution. In both the SEC and SV-AUC experiments, sample homogeneity manifests itself in the single and symmetrical peak. Furthermore, a native electrospray ionisation mass spectrometry experiment to check for the molecular weight of both the His_6-tagged wild type and H59A Sgt1 mutant shows that, while the latter is monomeric, it is also folded (as typically evidenced by a lack of peaks in the low mass:charge region of the product spectrum) (Figure 4.30).



Species	Stoichiometry	Expected Mass (Da)	Observed Mass (Da)	m/z Range
Sgt1 H59A	1	46957	46846.86 ± 8.91	3124.12 – 3904.91
	2	93914	93693.72 ± 17.82	4074.64 – 4932.25
	3	140871	140540.58 ± 26.73	5206.21 – 5622.62
Sgt1 D61R	1	47064	46997.57 ± 12.65	3134.17 – 3616.20
	2	94128	93995.14 ± 25.30	4087.75 – 4948.11
	3	141192	140992.70 ± 37.95	5222.95 – 5640.71

Figure 4.30: **Native Electrospray Ionisation Mass Spectrometry of Sgt1 Dimerisation Mutants – (A)** Native ESI-MS spectrum of Sgt1 H59A confirming the observation that H59A is a monomer in solution. **(B)** Native ESI-MS spectrum of Sgt1 D61R, showing the predominance of monomer and dimer species. Expected and observed masses are listed in the accompanying table for both sets of data. Spectra collection and data analysis was carried out by Richard Kerr.

Despite the presence of a solution-relevant dimerisation interface only a monomer of Sgt1 contacts Skp1 in the crystal structure, making the fundamental stoichiometry of an Sgt1:Skp1 complex 1:1. Therefore, in order to determine the role of dimerisation in the interaction with Skp1, a pulldown assay was performed using the dimerisation mutants H59A and D61R (Figure 4.31). The assay shows similar levels of Skp1 binding of the mutants and the wild type protein (Figure 4.31, Skp1 only, load versus bound). Despite the affect of these mutations on the solution behaviour of Sgt1 (Figure 4.29), and differences in the relative amounts of oligomeric species detected in the gas phase compared to the wild type (Figure 4.30), this does not appear to diminish Skp1 binding. Therefore, a dimer of Sgt1 is not required for Skp1 binding.

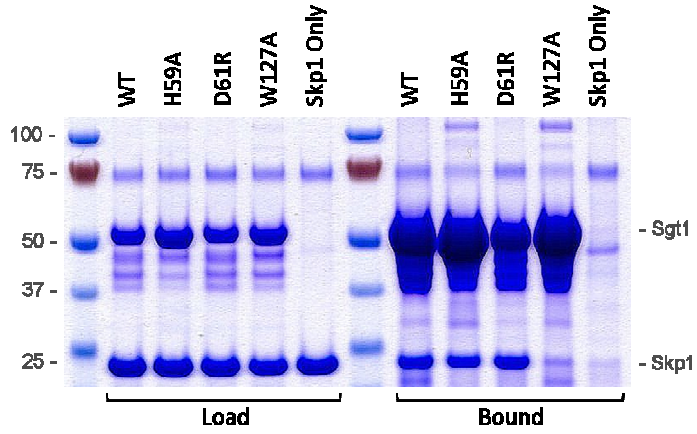
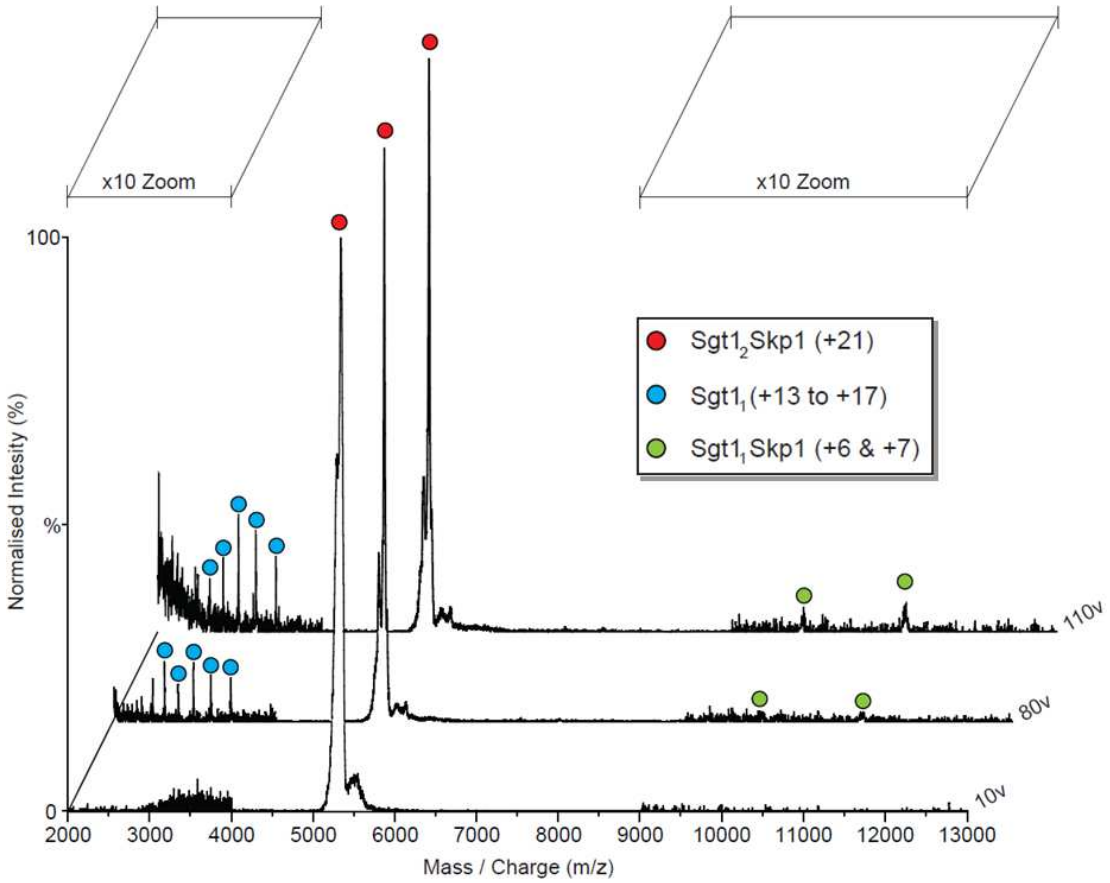


Figure 4.31: **Pulldown Assay of Sgt1 Mutants Probing the Importance of Dimerisation to Skp1 Binding –** Mutants affected in dimerisation (H59A and D61R) can bind Skp1 but to a lesser extent than the wild type. The lanes corresponding to the W127A mutant reiterates the importance of this region at the Sgt1:Skp1 interface and serves as an internal control for this pulldown.

This observation is supported by a dissociative mass spectrometry experiment on a quad-isolated 2:1 Sgt1:Skp1 complex peak, which showed that a monomer of Sgt1 dissociates before a 1:1 complex of Sgt1:Skp1. The strength of the Sgt1 to Skp1 interaction is therefore greater than the strength of the Sgt1 dimer (Figure 4.32).

However, the pulldown assay does not address the possibility that Skp1 binding compensates for the loss of inherent dimerisation of Sgt1 and promotes dimerisation of Sgt1 H59A. To assess this possibility, an analytical SEC experiment was carried out to compare the peak profiles of Sgt1 wild type and Sgt1 H59A in complex with Skp1 (Figure 4.33). It was expected that if Skp1

enforces dimerisation of Sgt1 (thereby justifying the 2:1 interaction), that the position of the peak centre of the H59A:Skp1 complex peak would overlap with the Sgt1 wild type peak. Any shift in the peak position of the Sgt1 proteins alone would indicate complex formation. Sgt1 WT and H59A were prepared at 50 µM concentration and Skp1 added in 2.7-fold molar excess. Fractions were collected across the peaks and analysed by SDS-PAGE to confirm the peak contents.



Species	Stoichiometry	Expected Mass (Da)	Observed Mass (Da)	m/z Range
Sgt1 FL:Skp1 FL	2:1	112742	115934 ± 7.36	5270.74 – 6102.80
	1:1	66856	69067.25 ± 5.70	9867.75 – 11512.21
Sgt1 FL	1	45886	46867 ± 1.66	3348.64 – 3906.58

Figure 4.32: **Tandem MS Study of an Sgt1:Skp1 Complex –** Shown here are the product spectra of an MS:MS study on an isolated Sgt1:Skp1 complex, with a voltage titration from 10-110 V. This data supports the relevance of the fundamental 1:1 interaction Sgt1:Skp1 seen in the crystal structure. Sgt1 monomers dissociate before a complex of Sgt1:Skp1 dissociates, suggesting that the Sgt1:Skp1 interaction is stronger than the Sgt1:Sgt1 interaction. Spectra collection and data analysis was carried out by Richard Kerr.

In the presence of Skp1, the peak corresponding to Sgt1 H59A alone shifts to a lower elution volume of 13.3 mL (Figure 4.33, A, magenta), denoting complex formation. A second peak at approximately 15.6 mL corresponds to excess Skp1

(Figure 4.33, A, pink). The SDS-PAGE analysis of fractions across the WT:Skp1 profile shows that the smaller peak centred at 12.3 mL contains both Sgt1 and Skp1. Therefore, both Sgt1 wild type and the H59A mutant bind to Skp1, which confirms the result from the pulldown assay (Figure 4.31). Furthermore, since the H59A:Skp1 peak does not coincide with the WT:Skp1 peak, the wild type interaction with Skp1 is probably 2:1, while the H59A:Skp1 interaction is 1:1.

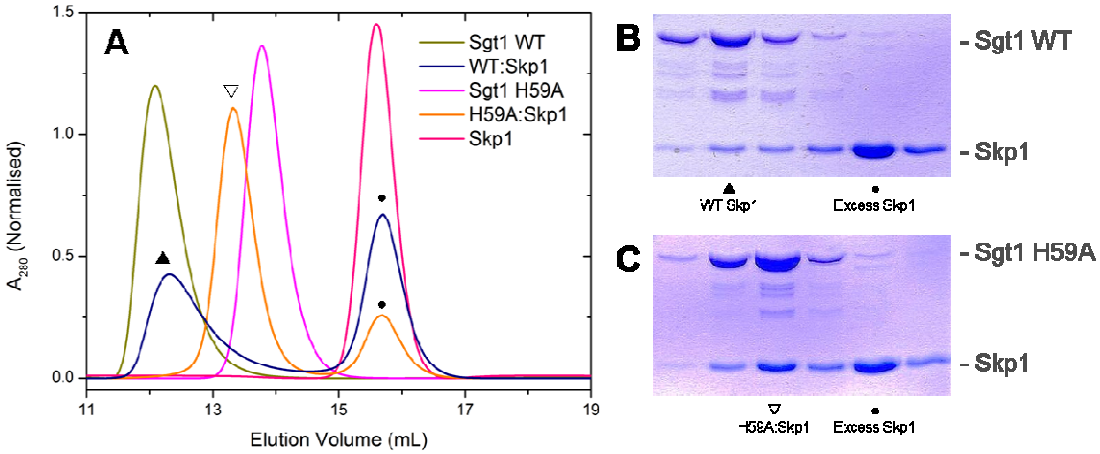


Figure 4.33: **Analytical Size Exclusion Chromatography of Sgt1 Wild Type and H59A in Complex with Skp1 – (A)** Wild type (WT) Sgt1 and H59A mutant Sgt1 were complexed with Skp1 to infer the stoichiometry of the resulting complex from a comparison of the peak shifts. WT Sgt1 alone (**Dark Yellow**) elutes at a lower volume compared to the H59A Sgt1 mutant (**Magenta**), which highlights the differences in their hydrodynamic behaviour. WT Sgt1 appears to form a complex with Skp1 (**Dark Blue**) (confirmed by SDS-PAGE analysis), which is presumably the stoichiometric 2:1 complex, while Sgt1 H59A mutant forms a 1:1 complex shown by the higher elution volume (**Orange**). The greater peak height of the excess Skp1 peak (at approximately 15.6 mL in both sets of complex sample peaks and Skp1 alone (**Pink**)) in the WT complex, is reflective of the smaller number of available binding sites for Skp1 on the Sgt1 dimer, by virtue of one of the sites being occluded. **(B)** SDS-PAGE analysis of fractions across the peaks of the Sgt1 WT:Skp1 complex. This analysis shows that the lower elution volume peak definitely corresponds to the Sgt1:Skp1 complex, since the lack of a peak shift does not confirm complex formation. **(C)** SDS-PAGE analysis of Sgt1 H59A:Skp1 complex peaks.

This is shown in the relative heights of the excess Skp1 peaks, which is greater for the wild type complex compared to the H59A complex (Figure 4.33, A, •, dark blue versus orange). While such a peak should be expected for both complexes, the differences in height reflects the relative number of binding sites accessible to Skp1 in solution on the wild type and mutant Sgt1 protein. The crystal structure shows that one of the concave binding grooves is occluded in the dimer conformation for Skp1 binding, which means that at equimolar starting concentrations, the wild type dimer has half the number of available sites compared to the monomerised H59A mutant. The fundamental interaction of

Sgt1 to Skp1 in solution is therefore 1:1, meaning that the function of a dimer of Sgt1 remains unresolved by these experiments.

4.9 Crystallisation of the Sgt1 TPR Domain

The hydrodynamic analysis of Sgt1 discussed in Chapter 3, suggests that Sgt1 might exhibit two modes of self-association. In order to determine whether this is the case, the Sgt1 TPR domain alone was crystallised for structural characterisation.

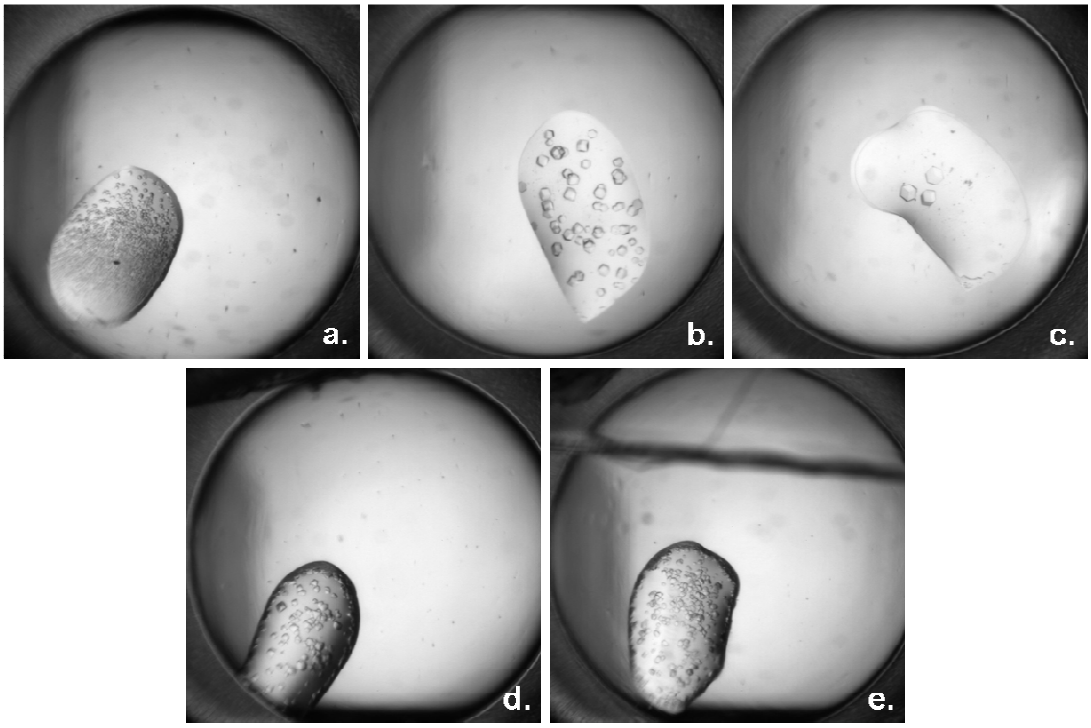


Figure 4.34: **Crystallisation Conditions for Sgt1 1-150 –** Crystals of Sgt1 1-150 were obtained at concentrations greater than 40 mg/ml and bore a cube-like morphology. Crystallisation attempts were conducted only under hanging drop conditions, with 150 nL total drop volume, dispensed using a Mosquito liquid handling robot. The initial hit **(a)** in 25% (w/v) PEG 3350, 200 mM $MgCl_2$ 100 mM Bis-Tris pH 5.5, yielded a shower of crystals, which were consequently optimised in a grid-screen testing PEG 3350 and $MgCl_2$ concentrations, giving larger crystals in smaller clusters in a condition with 27.5% (w/v) PEG 3350 and 280 mM MgCl2 **(b)**, and large cubic crystals in 30% (w/v) PEG 3350 and 300 mM $MgCl_2$ **(c)**. Additional crystals were obtained in 200 mM sodium malonate with 20% (w/v) PEG 3350 **(d)**, and 20% (w/v) PEG 6000, 200 mM MgCl2 and 100 mM Tris-HCl pH 8.0 **(e)**.

A broad screen of crystallisation conditions for Sgt1 1-150 yielded crystals of a cubic morphology in different conditions. An initial shower of crystals was optimised to yield large single cubes, which were subsequently harvested for

data collection at a synchrotron source (Figure 4.34). All consequent attempts at obtaining data for structure solution did not produce data to significant resolution: while crystal size was found to marginally affect the resolution of diffraction, data never extended beyond 6.0 Å.

In an effort to optimise conditions for structure solution of the Sgt1 TPR domain on its own, a new construct of Sgt1 based on its structure in complex with Skp1 (see previous discussion in this chapter) was cloned and expressed. This new truncation comprised the core TPR domain and capping helix, up to the last visible residue in the crystal structure (Asn136).

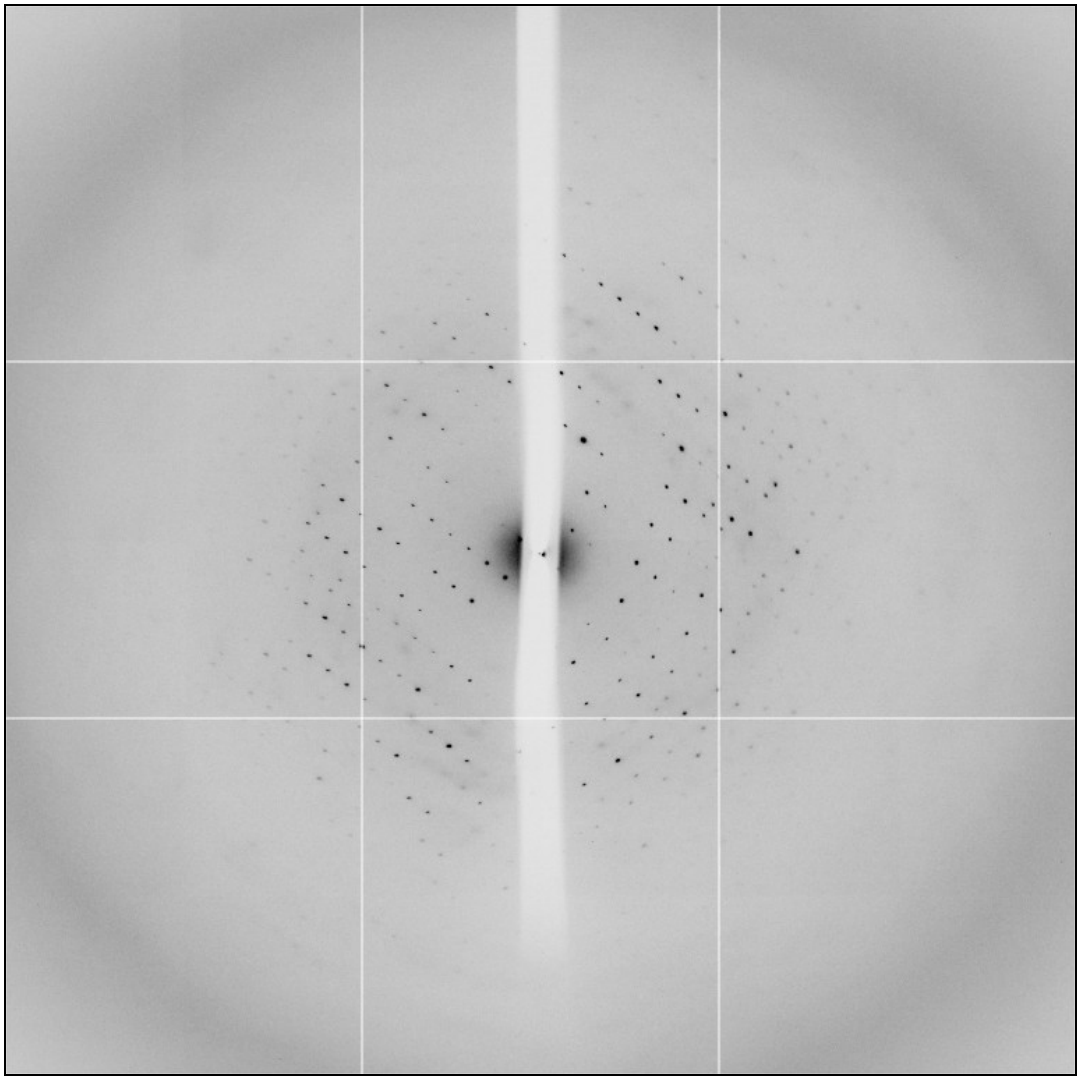
Crystals obtained of this optimised construct diffracted to approximately 4.3 Å at the Diamond synchrotron X-ray source (Figure 4.35). The data was automatically processed on-site and used in molecular replacement attempts in Phaser, despite the somewhat low resolution for this approach.

4.10 Structure Solution of the Sgt1 TPR Domain by Molecular Replacement

As the structure of the Sgt1 TPR domain had previously been solved in complex with Skp1, a single chain model of the domain was generated to be used for maximum-likelihood molecular replacement in Phaser. In the first instance, the whole TPR domain (including the C-terminal capping helix) was used as a model to place three or four copies of the TPR in the ASU (as suggested by the Matthews coefficient), with the expectation that the ASU arrangement of TPR protomers would be similar or identical to the Sgt1:Skp1 structure.

However, a Phaser query for three or four monomers produced only a partial solution with a LLG of 70 and refined translation Z-score of 11.5. While this suggested that Phaser had correctly placed one TPR protomer, the solvent content would have been unreasonably high (89.08%). In turn, different combinations of two TPR models, one with and one without the capping helix,

were trialled in various combinations in an attempt to phase the map (Figure 4.36, A).



	Overall	Inner	Outer			Overall	Inner	Outer
High resolution limit	4.31	13.63	4.31		Anomalous multiplicity	n/a		
Low resolution limit	61.24	61.24	4.54		Anomalous correlation	n/a		
Completeness	99.8	95.5	100.0		Anom. compl.	n/a		
Multiplicity	6.4	5.4	6.6		Anom. slope	n/a		
I/sigma	9.6	26.3	2.1		Partial bias	0.000	0.000	0.000
Rmerge	0.079	0.027	0.799		Total obs.	32733	999	4910
Rmeas (I)	0.093	0.032	0.940		Total unique	5112	184	745
Rmeas (I±)	0.091	0.031	0.941		Rpim (I)	0.048	0.017	0.485
Wilson B factor	0.000				Rpim (I±)	0.035	0.013	0.360

Unit Cell Dimensions	a (Å)	100.3	α (°)	90		Space Group	P 3_1 2 1
	b (Å)	100.3	β (°)	90			
	c (Å)	122.5	γ (°)	120			

Figure 4.35: **Diffraction Image and Statistics for a 4.3 Å Native Dataset from a Crystal of the Sgt1 TPR Domain**

The results of this search reflected that the C-terminal capping helix was positioned differently in at least one of the three (or four) monomers. Each Phaser query searching for a total of three monomers, including more than one copy of the cap-less TPR domain, produced single solutions with a final LLG of 461 and translational Z-score of 21.0. Searching for three copies of the cap-less TPR domain gave an LLG of 428 and translational Z-score of 19.9. No solutions were produced with four copies of the TPR domain, whether with or without the capping helix. The LLG for each search combination is summarised in Figure 4.36, B.

Upon inspection, each of the four solutions produced a similar starting map, with reasonable backbone density for the core TPR domain, but good density for the capping helix in only two of the three protomers. Where the capping helix was absent in the search model, positive F_O-F_C electron density was visible in the same position as the capped model, suggesting that it might pack in the same manner. In the third protomer, there appeared to be some additional positive F_O-F_C density main chain density for sequences beyond the capping helix, but directed away from the core TPR.

Instead of choosing a more complete asymmetric unit where one monomer already had the capping helix placed, the solution placing three copies of the cap-less TPR was chosen to begin model building. There were two main reasons for this. Firstly, the inability of Phaser to place two capped TPR protomers in the ASU suggested that there might be differences in the positioning of the capping helices. Secondly, building these from scratch would provide an opportunity to reinterpret their placement in a manner not biased by the packing of the capping helix in the search model. This was considered particularly important since one hypothesis, based on AUC and sequence comparison to CTPR3/CTPR3Y3 (see Chapter 3), was that Sgt1 might exhibit a mode of dimerisation where the capping helix is displaced or repositioned.

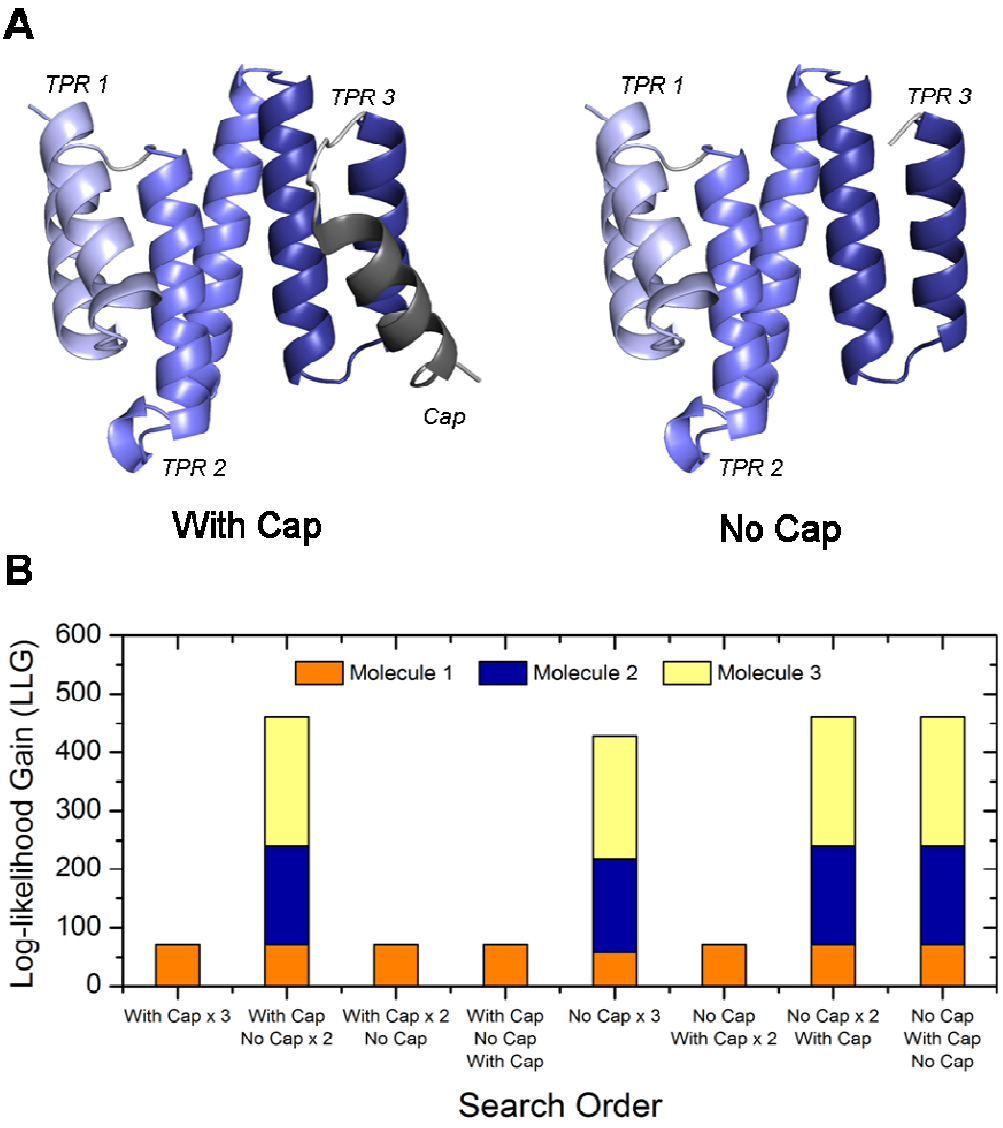


Figure 4.36: **Results of Phaser Molecular Replacement Search Combinations using a TPR Domain Model with and without the C-terminal Capping Helix – (A)** A single TPR chain from the previously solved Sgt1:Skp1 complex structure was used as a molecular replacement model in one of two configurations: with or without the C-terminal capping helix. **(B)** Summary of log-likelihood gain (LLG) of different search combinations. The Matthews coefficient suggested that there would be three copies of the TPR domain in the asymmetric unit. Different combinations searching for single or multiple copies of the TPR domain with and without the C-terminal capping helix were trialled. Searches were monitored by the accumulation of log-likehood gain scores. The result of this combinatorial search showed that the capping helix must be positioned differently in at least one of the three monomers, since a query searching for more than one copy of the TPR model with the capping helix yielded only partial solutions.

The molecular replacement model was re-built manually in Coot based on the visible electron density, and refined in Buster without TLS parameters or water placement to a final R_{Work} of 24.70% and R_{Free} of 26.25%. Both –autoncs and –target function were applied to improve the density on account of the low resolution of the dataset. The target model used was a single chain of the TPR

domain with a capping helix from the previously solved Sgt1:Skp1 structure. A Ramachandran analysis of the observed secondary structure reflects that the majority of residues are in favourable positions (99.22%) (Figure 4.37), with a final Molprobity score of 2.13 While this score is worse than the structure of the Sgt1:Skp1 complex, it is nonetheless in the 100th percentile of structures analysed by the Molprobity server (Chen VB *et al.*, 2010).

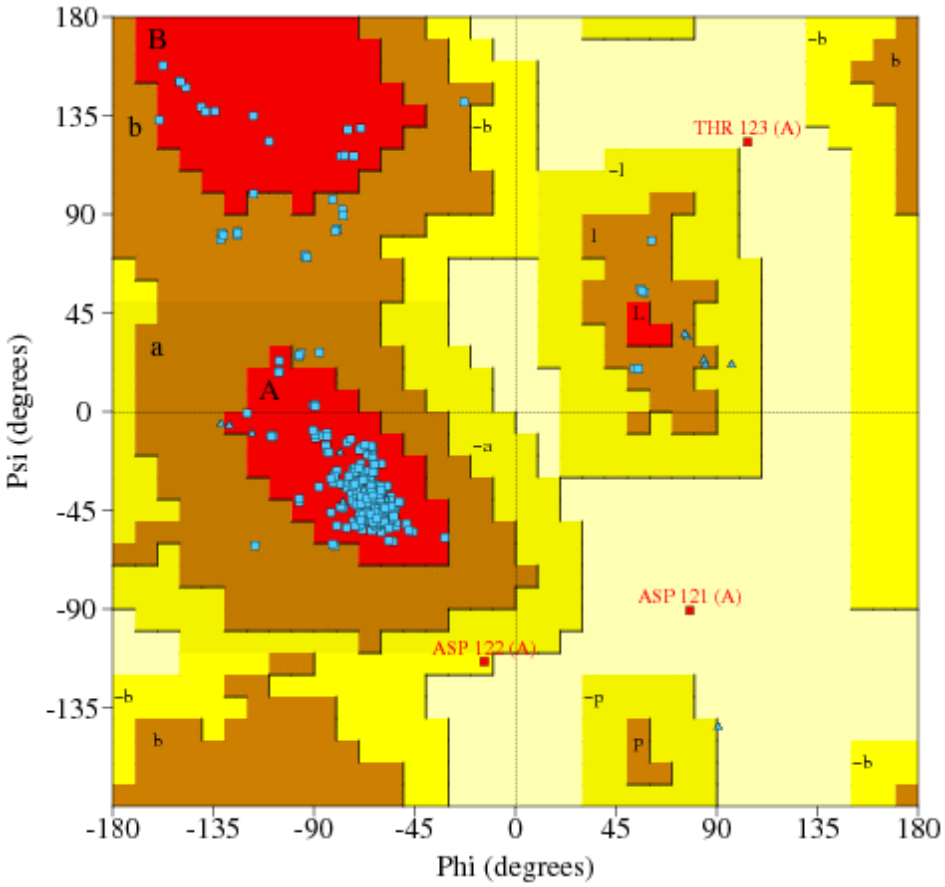


Figure 4.37: **Ramachandran Plot for the Structure of the Sgt1 TPR Domain –** After refinement in Buster, the geometry of the structure was analysed using the online server Molprobity (Chen VB *et al.*, 2010). 99.22% of all residues were in favourable regions, with only 0.52% Ramachandran outliers. In the final structure refined to an R_{Work} of 24.70% and R_{Free} of 26.25%. The final Molprobity score was 2.13, which is in the 100th percentile of macromolecular structures analysed using the server. This figure was generated using PDBsum (Laskowski, 2009).

The final refined structure of the Sgt1 TPR domain shows three copies of Sgt1 (chains A, B and C) in the ASU, with two of the TPR protomers (chains B and C) possessing a complete capping helix, while the last (chain A) has a partial C-terminus, indicative of some unravelling of the secondary structure in this region.

4.11 Composition of the Asymmetric Unit of a Structure of the Sgt1 TPR Domain at 4.3 Å Resolution

In contrast to the previously solved structure of the Sgt1 TPR domain in complex with Skp1, the TPR protomers in the ASU of this structure illustrate only one dimer conformation, while the other TPR:TPR contact is significantly smaller (Figure 4.38).

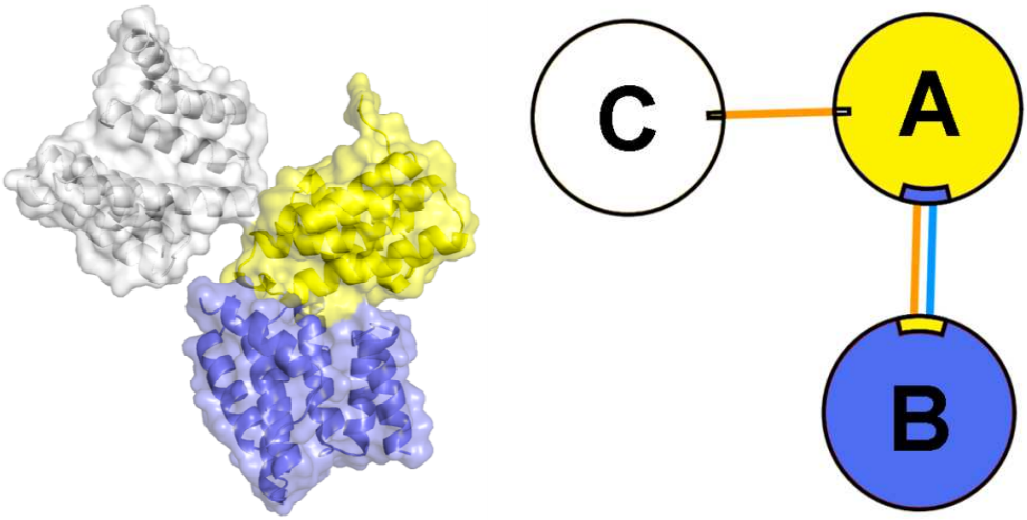


Figure 4.38: **Cartoon Representation of the Structure of the Sgt1 TPR Domain** – The structure of the Sgt1 TPR domain has three copies of the TPR domain in the asymmetric unit. The A:B interaction is identical to the primary dimer interface seen in the structure of the Sgt1:Skp1 complex, while the C:A interaction is a new but very small interface involving only a single hydrogen bond.

The electron density for each TPR protomer differs significantly. The backbone continuity within the core TPR domain is good for all three chains, though some of the loops and sidechains are poorly visible despite the application of NCS constraints and target structure refinement in Buster. Given that crystals of the Sgt1 1-150 construct never diffracted better than 6.0 Å, and 1-136 crystals only to 4.3 Å, the C-terminal region of the TPR domain probably has a significant amount of inherent flexibility. This notion is reinforced by the differences in the extent of visible electron density for the capping helix in each of the three protomers.

A direct comparison of the density of the region encompassing TPR helix 3B and the capping helix in each of the three chains, emphasises the differences between the three chains (Figure 4.39). The most consistent electron density is seen for chain C, where side chain density is visible for much of helix 3B and some parts of the capping helix. Chain C makes significant contacts with symmetry-related molecules, including an interaction with a symmetry-related chain B that is similar to the dimer interface involving the 3_{10} helix insertion. In turn, while the capping helix of chain B is probably fixed in position through this mode of dimerisation (via the stabilisation of TPR3), side chains in the capping helix of chain B remain solvent exposed due to a lack of contacts within and outside of the ASU.

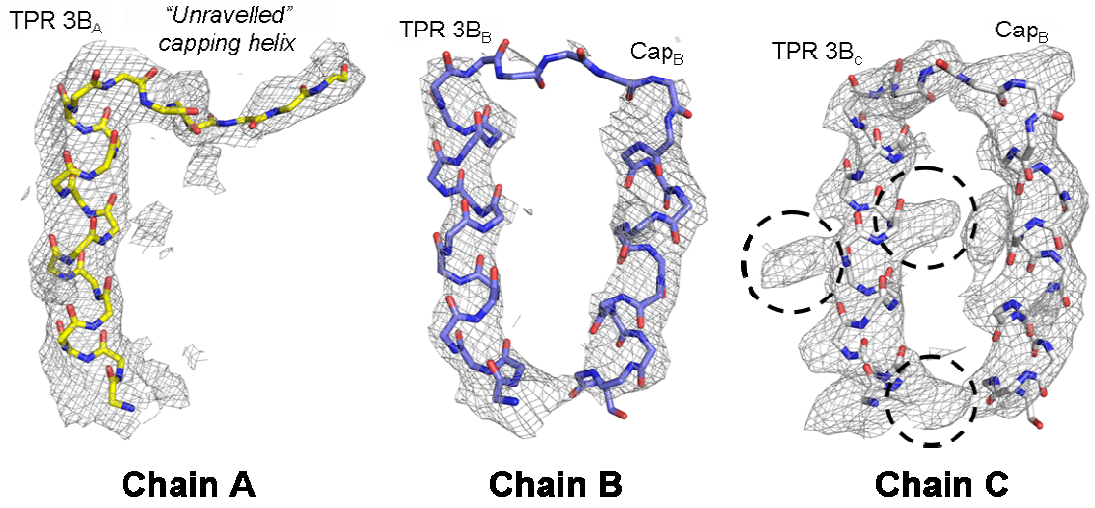


Figure 4.39: **Comparison of the Electron Density of Chains in the Asymmetric Unit in the Structure of the Sgt1 TPR Domain** – Differences in the three TPR chains is exemplified by the extent of continuity and completeness in the electron density of helix 3B and the capping helix. Similar to the Sgt1:Skp1 structure, chain A has the weakest density, chain B intermediate and chain C the most continuity in backbone density. In terms of side chain density, chain C is also the most resolved of the three, with some evidence for side chain density shown here (**Dashed Circles**).

This might account for the lack of sidechain density for helix 3B and the capping helix in chain B (Figure 4.39). A similar extent of completeness is visible for helix 3B in chain A, though because of the unravelled nature of the capping helix in this protomer, the density is quite poor. In comparison to the 2.8 Å map of the Sgt1:Skp1 complex, the electron density of the TPR domain in all three protomers is visibly poorer.

4.12 Evaluation of Sgt1 Dimerisation Interfaces Visible within the Crystal Structure

As eluded to previously, the asymmetric unit of the Sgt1 TPR domain on its own shows two sites of TPR:TPR contacts (Figure 4.40, A). The first mimics the dimer interaction that was previously shown to be relevant in solution (see Section 4.8), which involves the stacking interaction between His59 from one protomer (chain B) and Trp58 (chain A) from its neighbour (Figure 4.40, B).

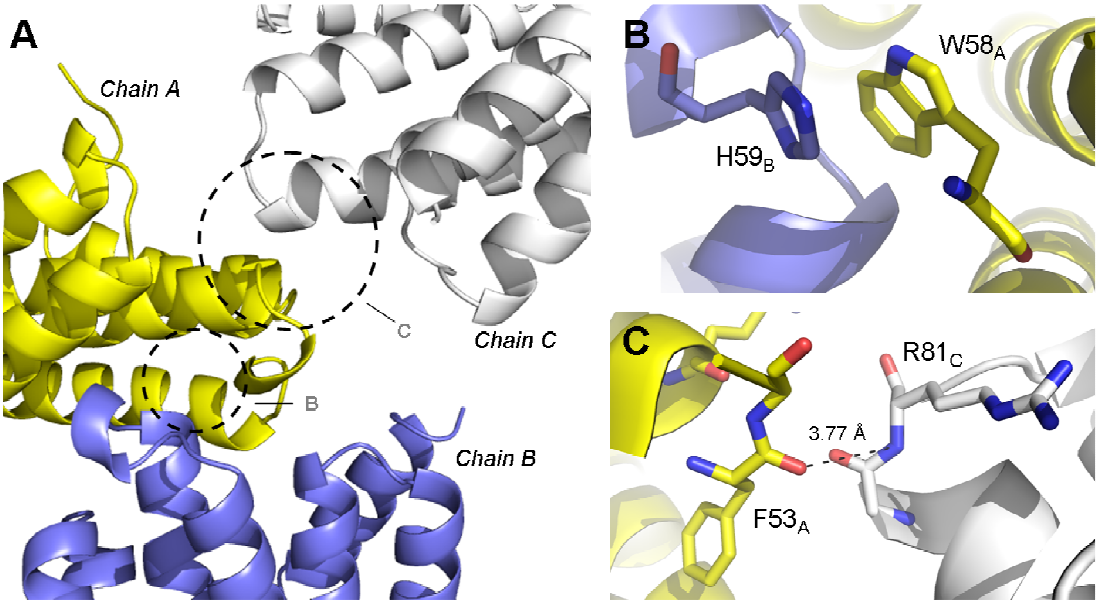


Figure 4.40: **Dimerisation Interfaces within the Asymmetric Unit of the Sgt1 TPR Domain Structure** – **(A)** The asymmetric unit of the structure of the Sgt1 TPR domain on its own shows two visible interfaces, one between chains A (**Yellow**) and B (**Blue**), and another between chains A and C (**White**). The A:B interface is identical to the previously characterised dimerisation interface, reflecting an interface that exists in solution, as shown by mutagenesis studies targeting this region. The second interface buries only 65.4 and 63.7 Å^2 of chains C and A respectively, with a single hydrogen bond between Phe53 of Chain A and Arg83 of Chain C. **(B)** The core of the main interface between chains A and B is a pi-stacking interaction between His59 from Chain B and Trp58 from Chain A. The interface buries 597.5 and 593.2 Å^2 of chains B and A respectively. **(C)** The lone hydrogen bond at the A:C interface is small and most likely not relevant in solution.

The interaction buries a total of 1190.7 Å^2 of surface area across the interface and shows the same core and peripheral interactions previously discussed (see Section 4.7). In addition, contacts of chain B in the ASU with a symmetry-related chain C molecule mean that the trimer arrangement seen in the Sgt1:Skp1 complex crystal structure is also present, but formed across two ASUs in this

crystal. The second interface is formulated by a small surface contact between chain A and chain C, burying only 129.1 Å^2 of surface area and including only a single hydrogen bond between main chain atoms of Phe53 on chain A and Arg81 on chain C (Figure 4.40, C). Since the interface is significantly smaller than other interfaces seen within the asymmetric unit, as well as contacts between ASUs, it is unlikely that this interface holds any relevance in solution.

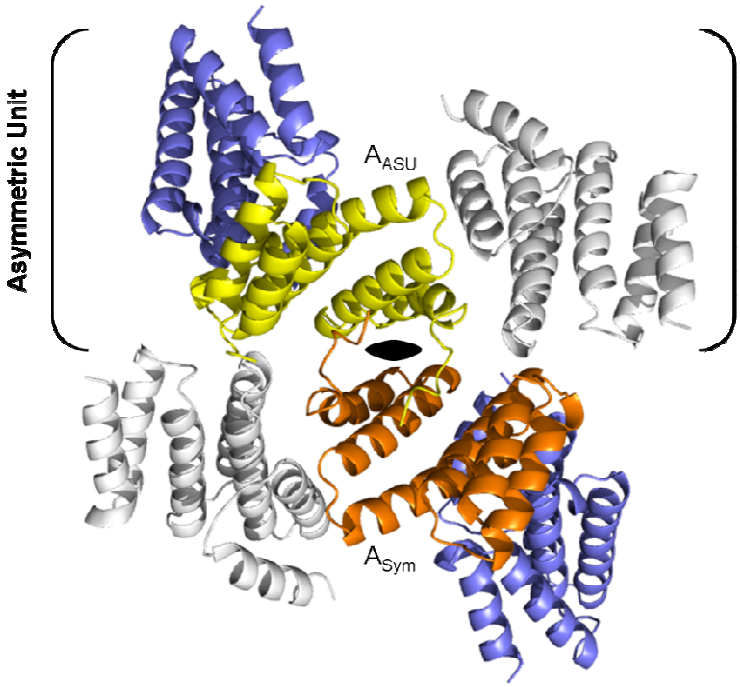


Figure 4.41: **Cartoon Representation of the New Interface Across a Two-Fold Symmetry Axis** – While the asymmetric unit contains only one significant dimer interface, an interface formed by two Chain A's, related by a two-fold symmetry axis contains an additional large Sgt1:Sgt1 interaction. Both TPR3 motifs of symmetry-related chain A TPR protomers are buried by this interface, and part of the unravelled capping helix is resolved in the concave groove of the TPR.

A closer inspection of possible interfaces using PDBe PISA (Krissinel *et al.*, 2007) shows that the largest interface is actually formed between two chain A molecules across a symmetry axis, which are related by the symmetry operation [–x, x+y,-z-1/3] (Figure 4.41). According to PISA, this interface also represents the most likely biological assembly, with a total buried surface area of 1570.1 Å^2 (785.0 Å^2 per monomer). The interface involves the exchange of the C-termini of adjacent protomers, in essence compensating for the exposure of hydrophobic residues that has occurred with the loss of structure within the capping helix (Figure 4.42, A).

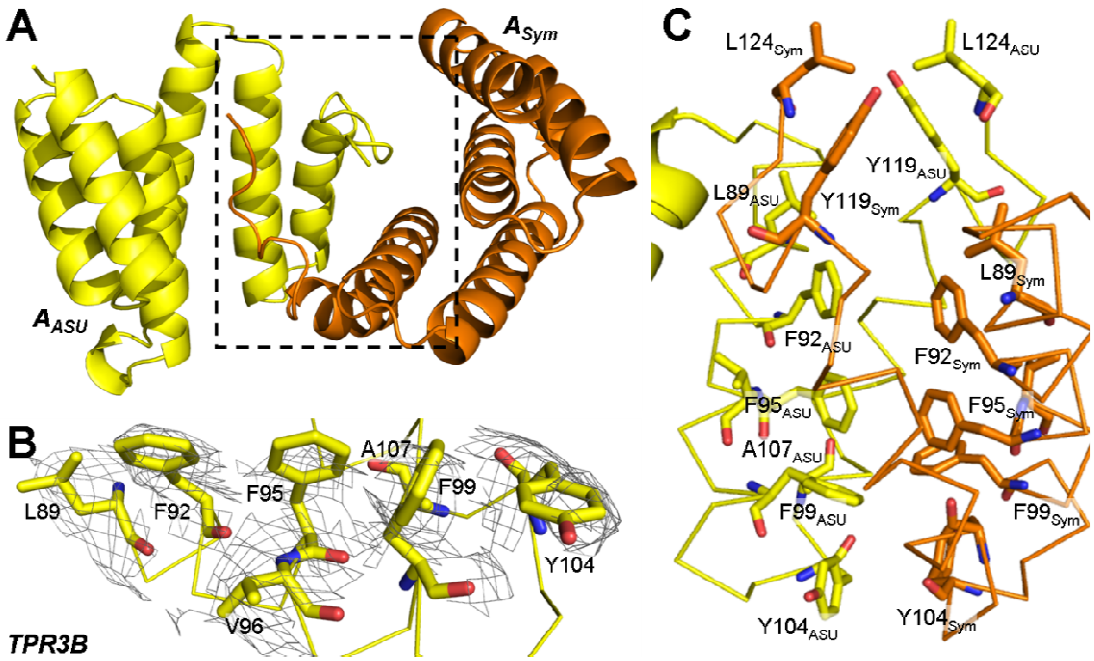


Figure 4.42: **Hydrophobic Core of the New Sgt1 TPR Dimer Interface** – **(A)** Cartoon representation of the quaternary structure of the new dimer interface. The interface involves the exchange of C-terminal sequences, which correspond to the capping helix. **(B)** $2F_O$-F_C electron density for residues in the core interface contoured at 1.48 sigma. **(C)** Summary of hydrophobic residues involved at the new interface shown in stick representation. While the interface has a few stabilising interactions (see Figure 4.41), it is predominantly hydrophobic, resulting in the burial of 1570.1 Å^2 of surface area across the interface.

The core interface involves an equivalent set of 21 hydrophobic residues in both interacting partners (Figure 4.42, C), with stabilising hydrogen bonds involving Tyr11, Asp121, Asp122 and Thr123 also mirrored across the interface (Figure 4.43). In addition, the conserved arginine residue (Arg93) at the centre of the concave TPR face forms a central salt bridge interaction with Asp121, in effect holding the extended conformation of the capping helix in place within the concave groove.

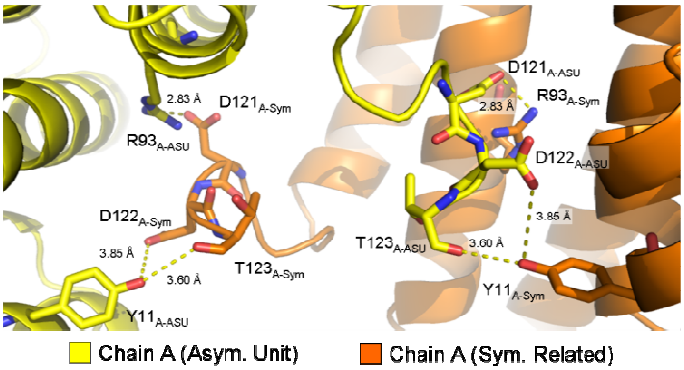


Figure 4.43: **Stabilising Interactions of the A_{ASU}:A_{Sym} Interaction** – The hydrophobic core of the new dimer interface is stabilised by hydrogen bonds including bonds contributed by Tyr11, Arg93, Asp121, Asp122 and Thr123 in both monomers. Chain A (A_{ASU}) is shown in **Yellow** and its symmetry-related counterpart (A_{Sym}) in **Orange**.

Analysis of the conservation of hydrophobic residues in yeast Sgt1 homologues shows that some of the well conserved residues are specific to this interface (Figure 4.44, A).

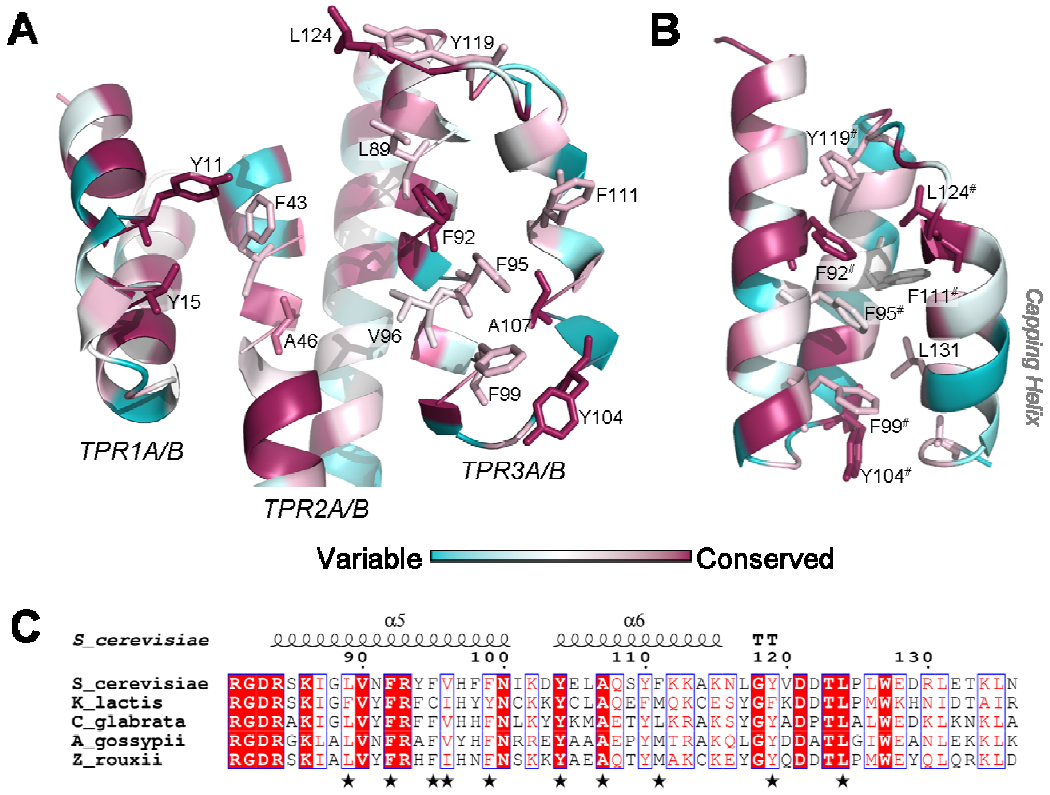


Figure 4.44: **Conservation of *S. cerevisiae* Sgt1 Residues Involved at the New Dimer Interface in Yeast Homologues of Sgt1** – **(A)** Cartoon representation of chain A of the structure of the Sgt1 TPR alone. Residues participating in the hydrophobic core of the new dimer interface are shown in stick representation, and are coloured by the degree of conservation in different species of yeast (the more intense the shade of magenta, the more conserved the position) as determined by the ConSurf server (Ashkenazy *et* al., 2010). **(B)** TPR3 and capping helix from a 'capped' TPR protomer in cartoon representation, with residues that mediate TPR3:capping helix packing shown as sticks. Residues are coloured by sequence similarity as in (A). Residues that overlap with the Sgt1:Sgt1 interface are marked by a #. **(C)** Sequence alignment of yeast Sgt1 homologues emphasising the conservation of TPR3 in yeast. The ostensibly less conserved residues (such as Leu89, Phe111 and Tyr119) are substituted for other hydrophobic residues in yeast, suggesting that this region is actually very similar amongst yeast orthologues of Sgt1. Residues that are involved in the 'closed'/cross protomer dimer interface are marked by « .

While Tyr11 and Tyr15 (and the conserved residues Asp122 and Thr123, not shown) are engaged at the Sgt1:Skp1 interface, the conserved residues Phe92, Tyr104 and Leu124 are unique to this interaction. None of the interacting residues participate in the previously characterised Sgt1:Sgt1 interaction

involving the conserved yeast loop region, which explains why both interactions appear in the crystal.

The sequence alignment further shows that some of the residues that are ostensibly less conserved are in fact substituted by equivalent residue types in other species of yeast (for example, Phe95 and Tyr119; Figure 4.44, C). Figure 4.40 shows that Phe92 and Phe95 sit at the centre of the hydrophobic core, while Tyr104 and Val119 flank this region on opposite ends of the interface. The high degree of sequence similarity amongst yeast homologues, and the fact that some of these residues are unique to this interaction, suggests that the interface might have some relevance in solution.

Contributing hydrophobic residues are positioned in TPR1 (Tyr11, Leu14, Phe43 and Ala46) and TPR3 (Leu89, Phe92, Phe95, Val96, Phe99, Tyr104, Ala107, Phe111 and Tyr119), as well as the region of the capping helix (Leu124) (Figure 4.44). These regions comprise the 'walls' of the concave groove of the TPR. Since the residues of the capping helix that are visible within the electron density lack any defined secondary structure, it is likely that the interactions being formed across this interface compensate for the lack of helical structure that would otherwise form in the presence of a binding partner in the concave groove (such as Skp1).

Residues in the interface overlap with those that mediate TPR3:capping helix packing in 'capped' TPR protomers (Figure 4.44, B). Therefore, differences in the positions of hydrophobic side chains in helices 3A and 3B, in comparison to the higher resolution TPR structure solved previously, might show that this new dimer conformation of chain A compensates for the instability of the capping helix. While the differences in resolution mean that indications of subtle changes in rotamer orientation or planar side chain movement in the structure of the TPR alone are less reliable, some positions (including Leu89, Leu117 and Tyr119) are significant enough for the electron density difference to be a real consequence of side chain movement in this region.

In chain C of the Sgt1:Skp1 complex structure, Tyr119 packs into the concave face of the core TPR domain. Rotational and planar shifts in the side chains of Leu89, Tyr94, Phe95, Phe111 and Leu117, propagating from the intra-TPR loop, up the helical axis of 3B in a C-terminal direction and leading into the 3B-capping helix loop (Figure 4.45), culminate in the repositioned Tyr119 at the top of helix 3B in chain A. In this conformation, Tyr119 is flipped out of the hydrophobic core that mediates packing of helices 3A and 3B against the capping helix and instead points in the direction of the C-terminus (Figure 4.45, $Y119_{No\ Cap}$ versus $Y119_{Cap}$). The propagation of side chain changes along the helical axis suggests that an 'uncapped' TPR domain is destabilised in the region C-terminal to the 3_{10} helix insertion. In turn, the unfavourable environment is stabilised by the inter-protomer exchange of sequences that are a result of the capping helix burying residues at the junction of helices 3A and 3B (Figure 4.42, C).

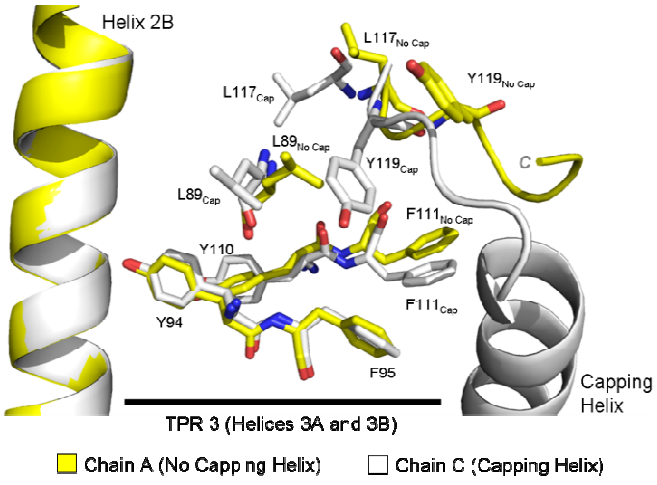


Figure 4.45: **Changes in the Network of Hydrophobic Interactions between Helices 3A, 3B and the Capping Helix** – Differences in the packing of hydrophobic residues in a 'capped' and 'cap-less' TPR protomer shows why the capping helix is displaced and unravelled in chain A, leading to the new dimer interface seen in the crystal structure of the Sgt1 TPR domain. The changes are localised to TPR3A and TPR3B and are propagated along the helical axis.

This cross-protomer interaction performs the same structural function as the TPR3:capping helix interaction in the other protomers, except that the large hydrophobic residues of helices 3A and 3B are interlaced with each other, rather than with those of the capping helix. It also fulfils the apparent need for a ligand within the concave groove of the TPR domain, since a number of the residues involved in the Sgt1:Skp1 interaction are also engaged in stabilising the unstructured C-terminal region (see previous discussion).

An interesting aspect of the Sgt1 TPR domain is the extended length of the loop region between TPR 3B and the capping helix, relative to the intra- and inter-TPR loop regions of the core domain (Figure 4.46). The intra-TPR loop between helices 1A and 1B is merely two residues in length and, excluding the 3_{10} helix insertion, helices 2A and 2B, and 3A and 3B, are each separated by only a single residue within the peptide backbone. In comparison, the distance between helix 3B and the capping helix is 5 residues, spanning Gly118 to Asp122 in those chains where the capping helix is fully resolved by the electron density. This is particularly relevant in Chain A, where the peptide backbone beyond Thr123 (the start of the capping helix) is directed in the opposite direction. The flexibility seems to stem from a hinge-like motion occurring around Gly118. Whether or not this extended conformation and flexibility is the cause of the rotation of the capping helix away from the TPR core is unclear, though it is definitely conducive to such a rotational movement.

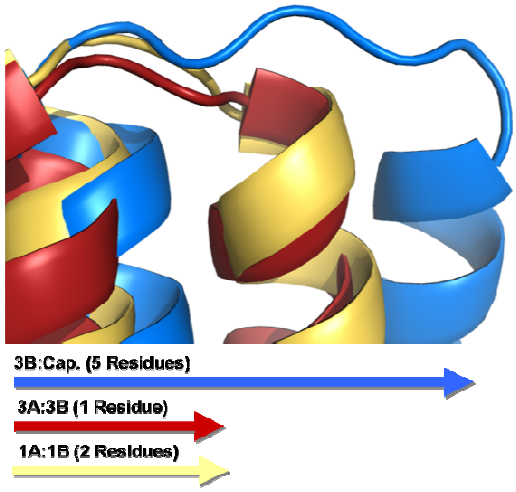


Figure 4.46: **Differences in Loop Lengths Between TPR Helices** – The increased length of the loop between TPR 3B and the capping helix is conducive to greater flexibility of the capping helix, which might be relevant in the absence of a ligand or binding partner within the concave groove of the TPR.

The structure presented here, of the TPR domain of Sgt1 alone, lends some support to the hypothesis of two dimer conformations of Sgt1, as originally inferred by biophysical characterisation of different Sgt1 constructs. Even though the resolution is poorer, it is sufficient to see major differences in side chain positions that result from or lead to the displacement of the capping helix in the core TPR. The biological relevance of this will be discussed further in Chapter 5.0.

CHAPTER 5: DISCUSSION AND CONCLUSION

5.1 BIOPHYSICAL CHARACTERISATION OF SGT1

Hydrodynamic studies of Sgt1 have re-emphasised its propensity for self-association and add further to the studies of Nyarko *et al*. and Bansal *et al*., which looked at the oligomerisation of Sgt1 in solution. Full length Sgt1 was shown by SV-AUC to exist predominantly as an elongated dimer (Figure 3.8), with a concentration-dependent propensity for higher order oligomerisation. This predominant behaviour is mimicked by a construct of Sgt1 truncated at the C-terminal SGS domain (Sgt1ΔSGS) (Figure 3.13), where the presence of a putative tetramer peak suggests that the SGS domain influences the formation of stable higher order oligomers of the full length protein. How the SGS domain contributes struturally to a dimer of Sgt1 is as yet undetermined, though the observations presented here are consistent with previous studies showing that the SGS domain negatively regulates dimerisation following phosphorylation by casein kinase 2 (Bansal *et al*, 2008), and suggest that it regulates the behaviour of Sgt1 *in vitro* (Figure 3.13).

Proof of trimerisaton in solution was obtained from experiments with a TPR construct of residues 1-150, which formed elongated trimers under the conditions of SV-AUC analysis (Figure 3.16). Native ESI-MS performed on 1-150 showed no monomeric species (Figure 3.15) and confirmed that these oligomers are stable in the concentration range examined. Therefore, sequences C-terminal to the core TPR modulate higher order self-association in a concentration-dependent manner.

Given the information in the existing literature, two complementary explanations for the distribution of the full length data were considered: a 'dead' equilibrium or two species in slow equilibrium. Both the 'dead' equilibrium and slow kinetics hypotheses imply that the published 20 nM K_D for a dimer of Sgt1 is inconsistent with the data presented here. The methods employed by Bansal *et al*. to

determine a nanomolar dissociation constant of the dimer lend little credence to their result, since this was derived using a micromolar concentration range of Sgt1 protein *in vitro*. To explore the possibility that the published K_D is incorrect, lower concentrations would be required in a similar experimental setup to the one used in this study. However, due to a low signal:noise ratio at such concentrations, a more direct method like isothermal titration calorimetry (ITC) might prove more suited to this instead.

Assuming a 20 nM K_D, the only consistent explanation therefore is higher order oligomerisation with fast kinetics. If this is the case, then domains C-terminal to the TPR domain (i.e. the CS and SGS domains) must be accommodated in these oligomers without physically hindering association of adjacent TPR domains. Attempts to explore a secondary self-association event failed due to non-ideal behaviour at concentrations greater than 22.3 μM (Figure 3.9). However, support for this idea comes from a combination of native ESI mass spectrometry experiments (Figure 3.9, 3.12 and 3.15), both of the crystal structures presented here (see Chapter 4 for further discussion) and SV-AUC of Sgt1 1-150 (Figure 3.16 and Figure 3.17) and 1-178 (Figure 3.18).

Solution studies of the Sgt1 TPR domain add two significant contributions to our current understanding of its function: firstly, that the TPR domain can form trimers, and secondly, that oligomerisation is moderated by sequences C-terminal to residue 150. The ability of Sgt1 1-150 to form trimers (Figure 3.16) led to the speculation that Sgt1 might exhibit two distinct modes of dimerisation, one of which becomes available upon the deletion of residues 151-178.

A bimodal frictional ratio model for SV-AUC of Sgt1 1-178 gave indications that two types of dimer conformations might co-exist in solution. In the two-peak distribution, the separation of species in terms of sedimentation coefficient represents a slow exchange of two dimer species in solution, where one species predominates over the other (Figure 3.17), and with one having an elongated and the other a more compact conformation. The molecular weight estimates (45516 Da for the main species, and 43550 Da for the minor species) suggest

that the distribution is the result of two separate pathways toward higher order oligomerisation, with both peaks representing an intermediate in a dimer:trimer equilibrium. At concentrations greater than 24.6 μM, the population is at a critical mass of components for one population to predominate over the other, and the exchange between the two is slow enough for them to be separated by SV-AUC. At the lowest concentration, the two species collapse into a single equilibrium. On a molecular level, this implies that two binding sites exist for Sgt1, with one having a higher affinity for a homomeric binding partner. This is supported by the Sgt1 TPR crystal structure, where two dimer interfaces were observed in the ASU (see Chapter 4).

It should be noted that, in light of this observation for TPR domain constructs, the differences in the main species distribution of full length Sgt1 before and after equilibration can be reconciled if two different subsets of dimers coexist in solution. One species is resolved when the system is not at equilibrium, with the shift in s-value reflecting fast kinetics, but the lack of peak broadening reflecting a stable equilibrium on the time scale of the experiment. After equilibration, the species converts into an equilibrium that maintains fast kinetics, but has a different equilibrium constant, as evidenced by a change in the peak profile with decreasing concentration.

Using SEC-MALS, the resulting trimer population of Sgt1 1-150 from the continuous c(s) distribution analysis was delineated into a combination of dimers and trimers, which together might account for the discrepancy in the molecular weight determined by AUC (Figure 3.21). The two sets of data complement each other by highlighting that Sgt1 1-150 undergoes primary and secondary steps in self-association (from monomer to dimer, and dimer to trimer), and that at any given concentration one population will predominate over the other (hence the s-value that is consistent with the predicted s-value for a dimer from HYDROPRO (Table 3.1)). It is important to bear in mind that these interactions occur only within the TPR domain, supporting previous data showing that this region is solely responsible for mediating self-association of Sgt1 (Bansal *et al*., 2008; Nyarko *et al*., 2007).

The delineation of solution equilibria is limited by the physical forces imposed on protein complexes in both AUC and SEC-MALS. In an SV-AUC analysis conducted at high speed, larger complexes are separated from smaller ones by the centrifugal force. However, if the rate of sedimentation does not discriminate between components and they are interchanging, then they will not be resolved. Conversely, in SEC-MALS, the dilution of complexes upon application to the column quickly removes smaller complexes away from an equilibrium, in turn changing the mass estimate for species in the MALS analysis. A non-equilibrium technique such as ESI-MS provides a snapshot of the equilibria in solution, and shows that the composition of Sgt1 and its truncations (at various concentrations) actually comprises a mixture of monomers, dimers and trimers (Figure 3.9, 3.12 and 3.15). This leads to the conclusion that Sgt1 exists in a dynamic equilibrium of monomers, dimers and trimers, and preferentially exists (at the concentrations tested) as a dimer, with some tendency toward trimerisation.

5.2 Structural Characterisation of Sgt1 Self-Association

The crystal structure of the Sgt1:Skp1 complex has shown that Sgt1 dimerisation can occur via a yeast-specific 3_{10} helix insertion, and involves a pi-stacking interaction between His59 on one protomer and Trp58 on the other (Figure 4.24). Self-associated TPR domains can bury a significant amount of surface area and commonly involve a number of helices in the TPR array. Cdc27 (PDB ID: 3KAE), for example, which forms part of the anaphase promoting complex, buries 2183.6 A^2 of surface area across its dimer interface and involves contributions from the N-terminal, middle and C-terminal regions of its six-repeat TPR motif (Zhang Z *et al*., 2010). The unique facet of the yeast-specific loop-mediated Sgt1 interface is that the interaction is rooted in a very specific region of the TPR domain, covering a total surface area of 1212.9 A^2 in the C:A dimer and 1257.1 A^2 in the equivalent A:B dimer. While this interface is relatively small compared to other three-repeat TPR dimers (e.g. Bub1, 1930.9 A^2), it compares favourably to a small subset of TPR dimer structures that were shown to exist in solution, including CTPR3Y3 (PDB ID: 2WQH) (1247.6 A^2 for

the end-on association and 1410.6 A^2 in the stacking-type interaction) and human type I collagen prolyl 4-hydroxylase (PDB ID: 1TJC) (1042.6 A^2) (Pekkala *et al*., 2004), and a single TPR:TPR interface in a trimer of YbgF (PDB ID: 2XEV) (1129.4 A^2).

The crystal structure of the isolated Sgt1 TPR domain has also brought to light a second interface that could hold some relevance for its function *in vivo*. The asymmetric unit of this crystal structure shows the previously discussed yeast-specific loop-mediated dimerisation interface (Figure 4.25 and Figure 4.40), and also reveals a larger interface between two symmetry-related TPR protomers (Figure 4.41). In this interface, an unravelled C-terminal capping helix is exchanged by an adjacent protomer, which results in the tight hydrophobic packing of adjacent TPR3 motifs and the burial of 1570.1 $Å^2$ of surface area. Here the concave binding surfaces of two TPR domains are 'closed' to additional quaternary interactions, and would require a rearrangement of Sgt1 into an 'open' arrangement to present the necessary sites for binding Skp1. While the SV-AUC analysis of the Sgt1 R93A mutant detailed in Section 4.9 shows that Arg93, despite forming a central salt bridge in this interface (Figure 4.43), is not important to dimerisation (Figure 4.29, right), its overall contribution to the interface is small compared to the hydrophobic core. Therefore, it is likely that targeting one of the key aromatic residues instead will be required to determine the importance of this dimer in solution.

While our structural studies provide little support for a role of dimerisation in binding Skp1 (see later discussion), dimerisation could be a necessity for stability of the N-terminal region. In this case, the reason that dimerisation defective mutants of Sgt1 (Bansal *et al*, 2008) result in kinetochore-specific defects stems from a compromised structure that makes Sgt1 incompetent for Skp1 binding. However, SEC analysis of Sgt1 dimerisation mutants H59A and D61R (and SV-AUC of H59A), which are affected at the yeast-specific interface (Figure 4.29, left), provide support for the stability of monomeric Sgt1 *in vitro*. In these experiments, the peak shape of Sgt1 H59A is similar to the wild-type. If dimerisation was important for the stability of Sgt1 and a monomer was

unfolded, then one might expect broadening of the peak profile or elution near the void volume, denoting aggregation. As this is not the case, dimerisation is not a measure for maintaining the fold of the N-terminus and the stability of the full length Sgt1 protein.

Two dimerisation modes have also been observed for CTPR3Y3 (Krachler *et al*., 2010) and SycD (Büttner *et al*., 2008). In SycD, a type III secretion chaperone, both an N-terminal head-to-head and head-to-tail interface are present in solution, but not critical to the stability of the TPR domain, as evidenced by SEC experiments on a double mutant affected at both interfaces simultaneously (Büttner *et al*., 2008). Being the mutant counterpart to CTPR3 (Main *et al.*, 2003), and therefore unaffected in the core TPR fold, a monomer of CTPR3Y3 would be expected to also be stable in solution in the absence of oligomerisation via either the end-on or loop-mediated stacking type interfaces. The monomer of H59A observed in SEC experiments therefore reflects a conformation where the capping helix is ordered and packing in a favourable position against the core TPR, but, because of the mutation, can no longer form dimers (Figure 5.3).

It should also be considered that the observed alternate dimer conformation of the TPR domain alone may represent a folding intermediate on the pathway to forming an 'open' conformation of the TPR domain with a structured capping helix. A similar folding intermediate was observed by Taylor and colleagues for the TPR domain of cyclophilin 40, a co-chaperone of Hsp90 that binds its EEVD motif and is involved in the regulation of steroid hormone receptor activity (Taylor *et al*., 2001). The two forms crystallise in different spacegroups, one monoclinic and one tetragonal. In the tetragonal form, sequences corresponding to TPR1 interface with TPR2A in a symmetry-related protomer (Figure 5.1). Residues C-terminal to the end of TPR2A (terminating in Glu298) are unstructured in this form. In the monoclinic crystal form, the TPR domain is fully folded and shows a three-repeat TPR fold with a capping helix. A comparison of the residues that are involved in TPR packing in the monoclinic and tetragonal form, showed that the packing of TPR helices across a symmetry axis mimics the TPR1:TPR2 interface in the folded TPR domain.

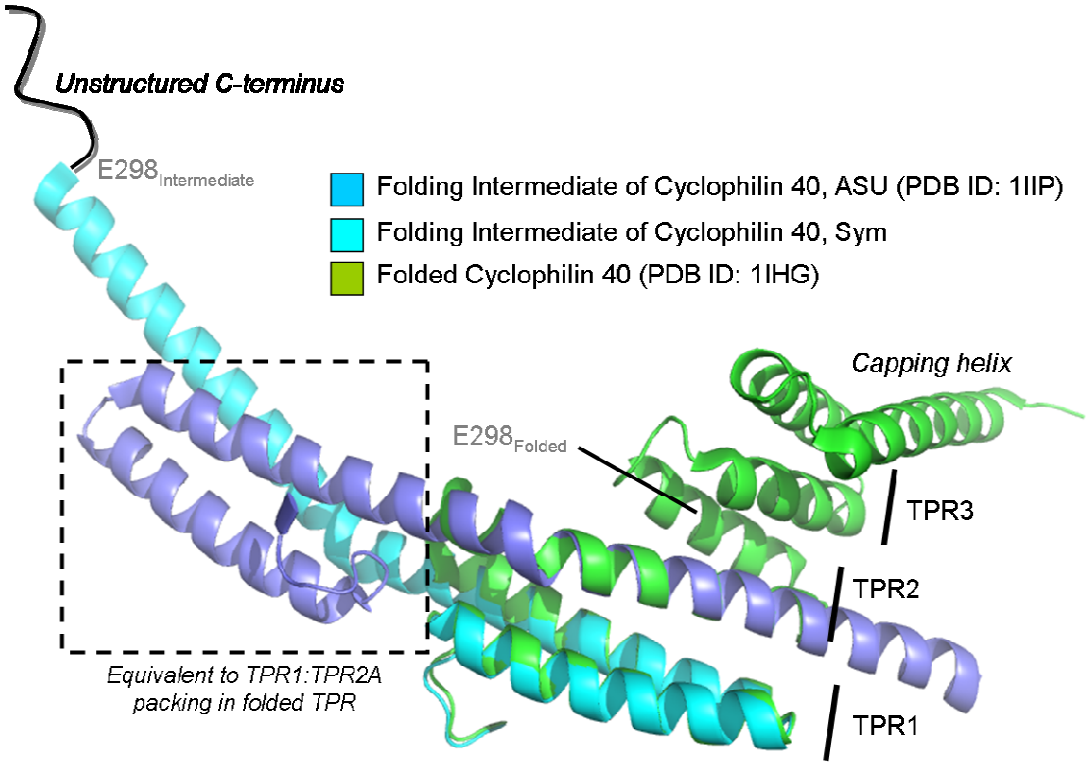


Figure 5.1: **Crystal Structures of the TPR domain of Cyclophilin 40 in Two Different Conformations** (PDB IDs: 1IIP and 1IHG) – Cyclophilin 40 crystallises in two crystal forms, one monoclinic and one tetragonal. In the monoclinic form, the TPR domain is fully folded and has three TPR repeats with a C-terminal capping helix. In the tetragonal form, only TPR1 and helix 2A in the TPR domain are ordered. An interaction occurs via TPR1 and TPR2A across two symmetry-related protomers that mimics the packing of TPR1 and TPR2 in the monoclinic form.

Analogously, in Sgt1, a number of the residues in TPR3 that are involved in packing the capping helix against the core TPR domain are also involved in similar cross-protomer TPR3 interactions. In the absence of structure in the capping helix, these interactions compensate for the unfavourable environment in the exposed helices of TPR3. The conformational state of the capping helix therefore determines the mode of dimerisation. The role of the capping helix in stabilising the TPR domain is exemplified by a truncation of the TPR domain to residue 121, which is insoluble when overexpressed in *E. coli* (data not shown).

In barley Sgt1, self-association is prevented by an intra-molecular disulphide bond that forms between helix 2A and the capping helix (Nyarko *et al.*, 2007), emphasising the role of the capping helix in oligomerisation. Residue types involved at this interface are conserved between yeast and plant orthologues (Figure 5.2). More importantly, however, one of the two cysteines that regulates dimerisation is in the region of intra-TPR packing of 3A and 3B, as shown by a

structure-based sequence alignment (Figure 5.2, yellow highlight). Disulphide linkage would therefore not only stabilise the capping helix secondary structure, but would also physically prevent this region from partaking in a quaternary interaction of two TPR protomers. It has been proposed that this disulphide bond acts as a sensor for the oxidative stress response to pathogen invasion in plants (Nyarko *et al*., 2007). Similarly in yeast, a potential conformational change from a 'closed' to an 'open' conformation might represent a switch mechanism that modulates the function of the N-terminus of Sgt1.

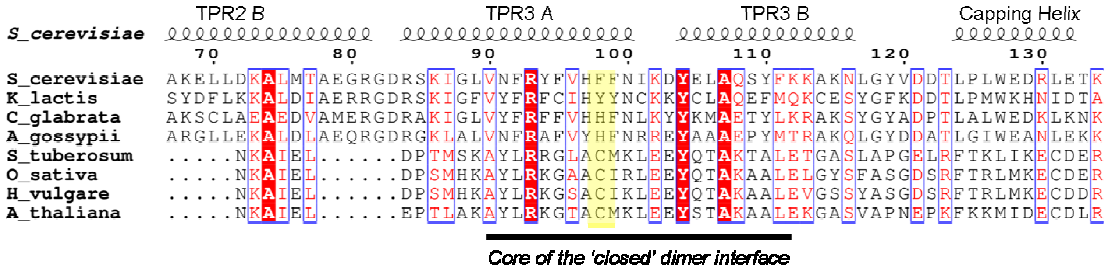


Figure 5.2: **Structure Based Sequence Alignment of Yeast and Plant Homologues of Sgt1** – This sequence alignment highlights the degree of conservation in TPR3 between yeast and plant orthologues of Sgt1. The presence of the functionally critical cysteine residue in TPR3A of plant Sgt1, lends support to the involvement of this region in a second dimer conformation of the Sgt1 TPR domain.

The fact that the structure of the Sgt1 TPR domain alone shows both the yeast-specific dimer and the cross protomer conformation highlights the possibility that they can exist simultaneously. This co-population of dimer species is shown in solution studies of Sgt1 1-178. If the 'closed' conformation is the first of two conformations sampled by Sgt1 after folding, then it must proceed through a series of steps to attain a stable capping helix (Figure 5.3).

In free solution and the absence of a ligand, Sgt1 exists in both the 'open' and 'closed' conformations, with an unstructured capping helix (Figure 5.3, Closed^{Unstructured} and Open^{Unstructured}). A fast equilibrium for the 'unstructured' conformations means that the capping helix is never in an unfavourable environment for long enough to cause aggregation, but also means that this conformation is not stable. Whenever this dimer is opened, the capping helix can passively sample a folding pathway leading to its restructuring (Figure 5.3, Open^{Structured}), which might proceed via a reorientation of Tyr119 followed by

Leu124 toward the hydrophobic core of the folded 3B:capping helix interface. In this way, given sufficient time for re-ordering *in vitro*, the capping helix restructures into a conformation that allows it to stabilise and pack against the core TPR domain. While there would be energetic barriers to this transition, the structured capping helix would be favoured to the unstructured form, making it unlikely that the species would revert to a previous point in the equilibrium. In SV-AUC of full length Sgt1, this conformation produces the shifting peak distribution after equilibrium, while before equilibration a part of the population is in the 'Unstructured' and the other in the 'Structured' equilibrium. At sufficiently high concentrations, TPR domain truncations of Sgt1 can be pushed to form the trimer seen in the crystal structure of the Sgt1:Skp1 complex.

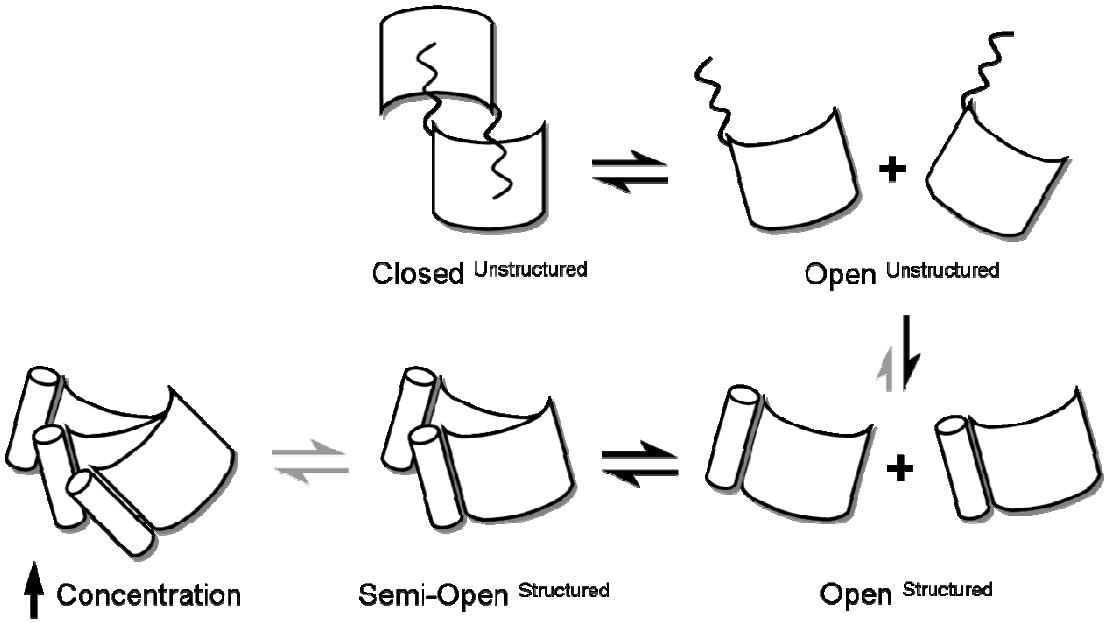


Figure 5.3: **Summary of the Possible Solution Equilbrium of Sgt1 –** The equilibrium of Sgt1 dimer conformations is perhaps more complex than a simple monomer:dimer equilibrium. The results detailed in this thesis exemplify different parts of the equilibrium, which are denoted in **Grey Text** underneath the respective steps. The **Superscript Text** describes the conformation of the capping helix.

5.3 Structural Characterisation of the Sgt1:Skp1 Interaction

Structural studies of the Sgt1:Skp1 interaction have revealed that Skp1 binds in the concave groove of the TPR domain (Figure 4.19). The region that mediates

this interaction has a high degree of sequence conservation in yeasts and higher eukaryotes, suggesting that this interaction is preserved amongst all Sgt1 orthologues. The interface involves a set of key hydrophobic residues on Sgt1, including Leu126 and Trp127, which form part of the capping helix of the TPR domain, and a conserved arginine residue (Arg93) that lies at the centre of the concave face (Figure 4.21). Mutating these residues to an alanine abolishes the interaction with Skp1, showing that, while the interface is small (1226.6 A^2), it reflects the relevant interaction in solution (Figure 4.28). Residues involved in the interaction are contributed by only one TPR protomer in a dimer of Sgt1.

This observation may be important when considering the stoichiometric interaction of Sgt1 and Skp1. In other TPR-mediated dimers such as the type III secretion system chaperones SycD (Schreiner *et al*., 2012) and IpgC (Lokareddy *et al*., 2010), a propensity for self-association is coupled to the need for distinct binding surfaces and consequent ligand stability; that is, the dimer conformations observed crystallographically are necessary for binding to and stabilising the cognate partners of both proteins. Conserved residues involved in the dimerisation of the yeast mitotic checkpoint protein Bub1 are also important for the interaction with the kinetochore protein blinkin (Bolanos-Garcia *et al*., 2009). In all three cases, TPR self-association is critical for function. Solution (Figure 3.22, 3.24 and 3.25) and gas phase studies (Figure 3.23) of various Sgt1:Skp1 complexes suggest that the biologically relevant stoichiometry is 2:1. However, the crystal structure suggests that only a monomer of Sgt1 is involved at the interface (Figure 4.19, A), where an Sgt1 dimer presents only one concave face, and masks the other through a concave-to-convex arrangement of protomers involved in the dimer. This is supported by mutational analyses of residues at the dimer interface (specifically, His59 to alanine), which abolishes dimer formation but maintains Skp1 binding (Figure 4.31 and 4.33). Therefore, in contrast to the role of dimerisation in SycD, IpgC and Bub1, dimerisation of Sgt1, while crucial to kinetochore formation (Bansal *et al*., 2008), does not appear to be required for its cognate partner Skp1. In this respect, dimerisation must therefore play some other role in initiating the assembly of the kinetochore.

Dimerisation may have arisen as a structural feature through the co-evolution of Sgt1 and Skp1. In this case, the surface area on an Sgt1 dimer is required specifically for stabilising some aspect of Skp1, such as the charged Skp1 loop. In our crystal structure, the loop region that is omitted in the Skp1Δ BTB/POZ construct is in position to make contacts with surface exposed regions occurring across the yeast-specific TPR dimer.

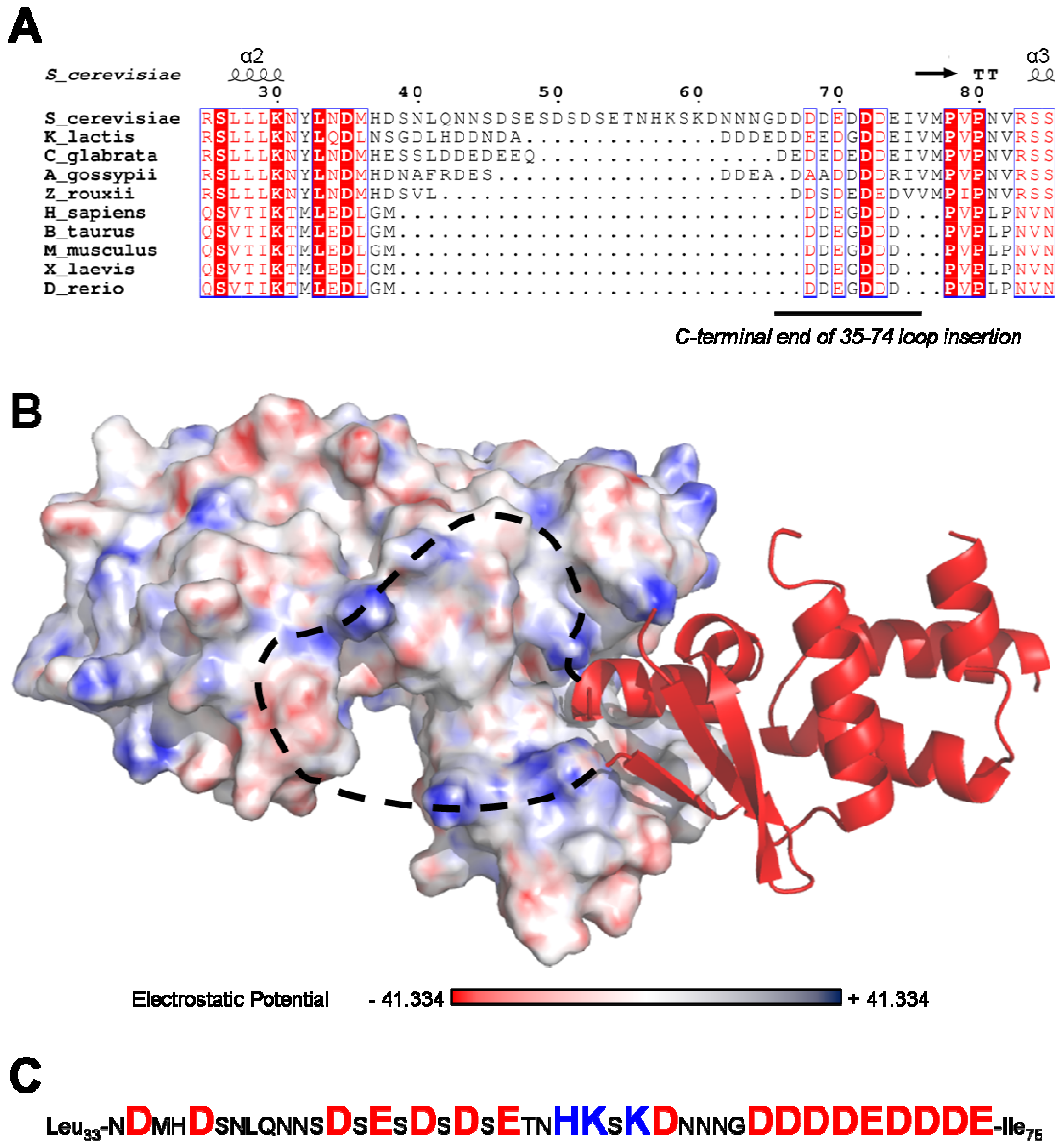


Figure 5.4: **The Conserved Loop Region of Skp1** – **(A)** Structure based sequence alignment using T-COFFEE/EXPRESSO (Notredame *et al.*, 2000) of Skp1 homologues shows conservation of the excised loop region in eukaryotes. The alignment image was made in ESPript (Gouet *et al.*, 1999). **(B)** Hypothetical positioning of the loop region, showing how the presence of the loop might stabilise both the Sgt1:Sgt1 interaction as well as the Sgt1:Skp1 interface. Two TPR protomers are shown in surface representation in the conformation seen in the crystal structure. Residues on Sgt1 are coloured by electrostatic potential. **(C)** Amino acid sequence of the excised loop region with negative (**Red**) and positive (**Blue**) residues represented in larger font.

There are two complementary features that support this model. First, a structure based sequence alignment of Skp1 sequences highlights the conservation of the C-terminal end of this region amongst all eukaryotes, hinting at a conserved function for this part of the loop (Figure 5.4, A). The charge distribution offered by a region of basic residues on an Sgt1 monomer could complement the numerous acidic residues found at the C-terminus of the Skp1 loop (Figure 5.4, B).

Second, the greater length of the loop region in yeast (in particular, *S. cerevisiae*) compared to higher eukaryotes means that the dimer accommodates the enhanced sequence divergence of yeast. This part of the loop also contains numerous amino acids with charged side chains (Figure 5.4, C). It is plausible that the extended length of this region in yeast might tether an Sgt1 dimer (Figure 5.4, B) and increase the apparent buried surface from 1219 $Å^2$ (discussed in Section 4.7). It is also possible that other regions – outside of the TPR domain – form additional interfaces with Skp1. Further support for this observation comes from co-purification attempts of Sgt1:Skp1 complexes, where increasing the ionic strength to greater than 200 mM sodium chloride hampers complex formation on gel filtration (data not shown). This suggests that electrostatic interactions make a significant contribution to the Sgt1:Skp1 interaction.

As a direct consequence of this, complex formation bridges together two TPR monomers, with the loop acting as a sort of 'molecular lasso' to stabilise the 2:1 Sgt1:Skp1 complex. This explanation is compatible with the observed 2:1 stoichiometry and maintains that the core of the interaction involves only *one* concave TPR binding surface as seen in the crystal structure. In higher eukaryotes, where the stoichiometry is 1:1 (Zhang *et al*., 2008), only a monomer of Sgt1 is required for the functional Sgt1:Skp1 interaction because this loop region is significantly shorter compared to its yeast counterpart (Figure 5.4, A).

In this respect, Sgt1 is similar to PP5, which exhibits conformational flexibility in the absence of a ligand in the concave groove, but stabilises in the presence of an Hsp90 MEEVD peptide (Cliff *et al*., 2005). In order to stabilise the loop region of

Skp1, a second TPR protomer is required to bind via the conserved 3_{10} helix insertion in yeast, which presents additional binding sites for the Skp1 loop. In the presence of a ligand, Sgt1 therefore has a three stage conformational cycle, similar to those described previously for no ligand: a stable 'closed' conformation, a metastable 'open' conformation where the capping helix remains unstructured but the concave groove is primed for ligand binding, and a stable 'open' conformation, for ligand stabilisation.

The fundamental differences in the structuring of the C-terminal capping helix in the two Sgt1 dimer conformations might explain why they do not exist simultaneously in the presence of Skp1, and why a 3:1 Sgt1:Skp1 complex does not form in solution (Figure 3.25). Allowing a mixed population of Sgt1 and Skp1 to incubate for a prolonged period would provide the timescale necessary for one interface to re-order into a conformation that is compatible with Skp1 binding. If Skp1 samples the 'open' conformation, then this facilitates proper folding of the capping helix to stabilise its binding in the concave groove. In this conformation, Sgt1 does not readily form trimers in the concentration range examined, even when Sgt1 is in significant molar excess relative to Skp1, because the capping helix is preferentially ordered and therefore does not participate in an interface across two TPR protomers. Even though the crystal structure shows a trimer of Sgt1 and a 3:1 stoichiometry of Sgt1 to Skp1, the concentration of components in the crystal exceeds those that would be found in the cell. These high concentrations therefore push the equilibrium toward the observed trimer conformation. In the full length protein, the C-terminal CS and SGS domains (which are absent in the crystallised construct) would prevent this type of trimerisation. Therefore, the 3:1 ratio of Sgt1:Skp1 is an artifact of crystallisation and not relevant in solution.

5.4 Implications for the Role of Sgt1 in Kinetochore Assembly

Kinetochore assembly is dependent on Sgt1 and Hsp90 (Bansal *et al*., 2004; Linglebach *et* al., 2004; Steensgaard *et* al., 2004). To date, there is little understanding for the role of Hsp90 in this context, other than the fact that it is

not required for dimerisation of Sgt1 (Bansal *et al.*, 2008). While Hsp90 stimulates the interaction of Sgt1 and Skp1 3.2-fold (Bansal *et al.*, 2004), our own experiments have shown that a complex of Sgt1 and Skp1 can from *in vitro* in the absence of Hsp90. Preliminary experiments suggest that co-purification of all three proteins leads ultimately to the separation of an Hsp90 dimer and an Sgt1:Skp1 complex in size exclusion (data not shown) reflecting the metastable nature of this heteropentamer *in vitro*.

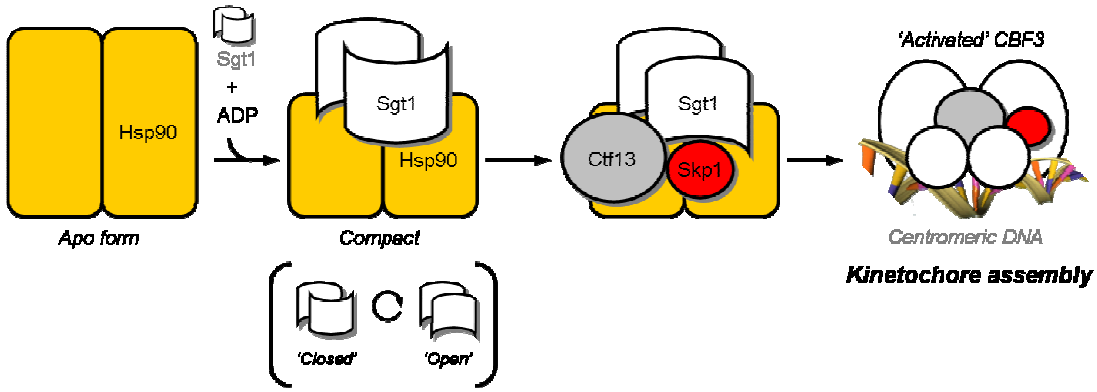


Figure 5.5: **An Updated Model for Kinetochore Assembly** – The data presented herein point to a specific role for Hsp90 in the assembly of the yeast kinetochore in stabilising a structure of the Sgt1 N-terminus that is compatible with binding the CBF3 component Skp1. In the compact ADP-bound conformation, Hsp90 also provides a physical scaffold for Sgt1-mediated bridging of the Skp1:Ctf13 heterodimer that is required for initiation of kinetochore assembly.

The current model for kinetochore assembly lacks a significant amount of detail on the role of Hsp90 in this pathway. Hsp90 exhibits multiple conformational states that are conducive to client protein binding and activation (discussed in Chapter 1), with Sgt1 preferring the lid-open ADP-bound conformation (Catlett *et al.*, 2006). Inhibition of Hsp90 function by the specific inhibitor 17-AAG causes the improper localisation of kinetochore proteins in human cell lines (Niikura *et al.*, 2006), and overexpression of Hsp90 rescues kinetochore defect mutants of yeast Sgt1 (Bansal *et al.*, 2004). In yeast, the Sgt1:Skp1 interaction via Hsp90 is important for the activation of Ctf13, the structural core of the CBF3 DNA binding anchor of the kinetochore (Stemmann *et al.*, 2000). In humans, Hsp90 and Sgt1 regulate chromosome alignment through Mis12, a complex that regulates kinetochore-microtubule attachment. Co-inhbition of Sgt1 and Skp1 results in the stability of Mis12 at the kinetochore, suggesting that Mis12 complexes are regulated via the ubiquitin ligase pathway (Davies *et al.*, 2010). Hsp90 is

therefore functionally implicated in kinetochore regulation across the eukaryotic domain, but its mechanistic contribution is less clear. The data presented in this thesis allows us to formulate an updated model for kinetochore assembly, bearing in mind the unknown mechanism of Hsp90 for this function (Figure 5.5) and that the second dimerisation interface in the Sgt1 TPR structure alone has not yet been validated.

Instead of performing an active role, Hsp90 passively stimulates the dissociation of the Sgt1 dimer into the metastable 'open' state, by providing a favourable environment for the folding of the Sgt1 capping helix in its compact ADP-bound conformation. This would account for the stimulation of Sgt1:Skp1 binding in the presence of Hsp90, since Hsp90-binding would inherently favour the Sgt1 conformation relevant to this interaction. In addition, it also explains why the kinetochore defect phenotype of the yeast *sgt1-3* mutant, which has a mutation at Phe99 (one of the key residues involved in 'closed' dimerisation), is suppressed upon overexpression of Hsp90 (Bansal *et al.*, 2004). Binding of Skp1 to the TPR domain re-engages the TPR dimer via the steps described previously. Once this has occurred, the scaffolding function of Hsp90 is no longer required, since all components have been stabilised, which is reflected in the relative affinity of components of the heteropentamer in co-purification by size exclusion chromatography.

Hsp90, in this capacity, is involved in the folding of Sgt1. Since Ctf13 is believed to be a client of Hsp90 (Stemmann *et al.*, 2002), the Sgt1-coupled compact ADP-bound conformation would provide the physical bridge between Skp1 and Ctf13, following Hsp90-mediated folding of Sgt1, and Sgt1-mediated presentation of Skp1 to Ctf13. The fact that the SGS domain binds leucine rich repeat domains (Austin *et al.*, 2002) supports the notion that Sgt1 serves as an adaptor for this interaction. In this respect, the conformational switch at the N-terminus of Sgt1 acts as a regulator for the activation of the Skp1:Ctf13 heterodimer, and the forward transition from an unstructured to a structured capping helix is stimulated by Hsp90.

5.5 FUTURE WORK

The work presented in this thesis has answered questions related to the structural basis for the Sgt1:Skp1 interaction, as well as the self-association of Sgt1, but has in turn unveiled more questions regarding the mechanistic consequences for kinetochore assembly.

In the first instance, it will be necessary to uncover the importance of the Skp1 loop region to the Sgt1 interaction and perhaps probe whether the sequence diverge between yeast and higher eukaryotes points to a co-evolution of Sgt1 and Skp1. This can be addressed by isothermal titration calorimetry (ITC) of Skp1 with and without the loop, titrated into a solution of full length Sgt1. In addition, given the importance of the second interface to the working hypothesis, mutagenesis experiments targeting the 'closed' conformation would need to be conducted and the oligomeric state of Sgt1 investigated by one of the solution methods utilised in this study. A good starting point is the F99L mutation of the *sgt1-3* mutant, or a more severe mutation to an alanine, to truly test the significance of this interface in solution.

Secondly, the proposed role of Hsp90 involves a very significant conformational rearrangement in Sgt1. This can be probed by biophysical techniques including electron paramagnetic resonance (EPR) spectroscopy or Förster resonance energy transfer (FRET)-based experiments. Both techniques could monitor the distances between labels positioned in the 3_{10} helix insertion with and without Hsp90, to understand whether or not a conformational change is facilitated when Sgt1 is complexed to the ADP-bound form of the chaperone. Alternatively, FRET labels could be positioned at the top and end of the capping helix, as well as the start of the CS domain, to probe distance changes between these domains upon Hsp90 addition. The rationale for this experiment would be that stimulated ordering of the capping helix in the presence of Hsp90, would bring the three labels into closer proximity for FRET. Cross-linking experiments, stabilising one of the conformations of Sgt1, would complement this approach by showing the relative accessibility of binding sites for Skp1.

If the proposed hypothesis is correct, then Sgt1, Skp1 and Ctf13 could form a stable complex for crystallisation. Solving this structure would prove (or disprove) that Sgt1 acts as an adaptor protein similar to Hop, and is critical in activation the structural core of the CBF3 complex. Coupled to an investigation of the effect of Hsp90 on the conformation of the Sgt1 N-terminus, this would add further information to the role of Hsp90 in the kinetochore assembly pathway.

By starting with this set of experiments, it should be possible to gain insight into the dynamic role of Hsp90 and its biology. However, the focus of these experiments remains on Sgt1, the unique co-chaperone molecule that ties together all of the pathways relevant to kinetochore assembly in yeasts.

~

REFERENCES

1. Abe Y, Shodai T, Muto T, Mihara K, Torii H, Nishikawa S, Endo T, Kohda D. 2000. Structural basis of presequence recognition by the mitochondrial protein import receptor Tom20. Cell. 100 (5): 551-60

2. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R,Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010. PHENIX: a comprehensive Python-based system for macrmolecular structure solution. Acta Crystallogr D Biol Crystallogr. 66 (Pt. 2): 213-221.

3. Ali MM, Roe SM, Vaughan CK, Meyer P, Panaretou B, Piper PW, Prodromou C, Pearl LH. 2006. Crystal structure of an Hsp90-nucleotide-p23/Sba1 closed chaperone complex. Nature. 440 (20): 1013-1017.

4. Allan RD, Ratajczak T. 2011. Versatile TPR domains accommodate different modes of target protein recognition and function. Cell Stress Chaperones. 16 (4): 353-367.

5. Andersen RO, Turnbull DW, Johnson EA, Doe CQ. 2012. Sgt1 acts via an LBK1/AMPK pathway to establish cortical polarity in larval neuroblasts. Dev Biol. 361 (1): 258-265.

6. Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics. 22 (2): 195-201.

7. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. 2010. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucl Acids Res. 38 (Web Server issue): W529-W533.

8. Austin MJ, Muskett P, Kahn K, Feys BJ, Jones JD, Parker JE. 2002. Regulatory role of *SGT1* in early *R* gene-mediated plant defenses. Science. 295 (5562): 2077-2080.

9. Azevedo C, Betsuyaku S, Peart J, Takahashi A, Noël L, Sadanandom A, Casais C, Parker J, Shirasu K. 2006. Role of SGT1 in resistance protein accumulation in plant immunity. EMBO J. 25 (9): 2007-2016.

10. Bai C, Sen P, Hofmann K, Ma L, Goebl M, Harper JW, Elledge SJ. 1996. SKP1 Connects Cell Cycle Regulators to the Ubiquitin Proteolysis Machinery through a Novel Motif, the F-Box. Cell. 86 (2): 263-274.

11. Bansal PK, Abdulle R, Kitagawa K. 2004. Sgt1 Associates with Hsp90: an Initial Step of Assembly of the Core Kinetochore Complex. Mol Cell Biol. 24 (18): 8069-8079.

12. Bansal PK, Mishra A, High AA, Abdulle R, Kitagawa K. 2009. Sgt1 dimerisation is negatively regulated by protein kinase CK2-mediated phosphorylation at Ser361. J Biol Chem. 284 (28): 18692-8.

13. Bansal PK, Nourse A, Abdulle R, Kitagawa K. 2009. Sgt1 dimerisation is required for yeast kinetochore assembly. J Biol Chem. 285 (6): 3586-92.

14. Barford D. 2011. Structural insights into the anaphase-promoting complex function and mechanism. Philos Trans R Lond B Biol Sci. 366 (1584): 3605-3624.

15. Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG. 2011. *iMOSFLM*: a new graphical interface for diffraction-image processing with *MOSFLM*. Acta Crystallogr D Biol Crystallogr. 67 (Pt 4): 271-281.

16. Bellizzi JJ, Sorger PK, Harrison SC. 2007. Crystal Structure of the Yeast Inner Kinetochore Subunit Cep3p. Structure. 15 (11): 1422-1430.

17. Benesch FLP, Ruotolo BT, Simmons DA, Robinson CV. 2007. Protein Complexes in the Gas Phase: Technology for Structural Genomics and Proteomics. Chem Rev. 107 (8): 3544-3567.

18. Bergerat A, de Massy B, Gadelle D, Valoutas PC, Nicolas A, Forterre P. 1997. An atypical topoisomerase II from Archaea with implications for meiotic recombination. Nature. 386 (6623): 414-417.

19. Biggins S, Severin FF, Bhalla M, Sassoon I, Hyman AA, Murray AW. 1999. The conserved protein kinase Ipl1 regulates microtubule binding to kinetochores in yeast. Genes Dev. 13 (5): 532-544.

20. Bhattacharya S, Lee YT, Michowski W, Jastrzebska B, Filipek A, Kuznicki J, Chazin WJ. 2005. The modular architecture of SIP facilitates its role in stabilizing multiprotein assemblies. Biochemistry. 44 (27): 9462-9471.

21. Black BE, Foltz DE, Chakravarthy S, Luger K, Woods VL Jr, Cleveland DW. 2004. Structural determinants for generating centromeric chromatin. Nature. 430 (6999): 578-582.

22. Bloustine J, Berejnov V, Fraden S. 2003. Measurements of Protein-Protein Interactions by Size Exclusion Chromatography. Biophysical Journal. 85 (4): 2619-2623.

23. Bohen SP. 1995. Hsp90 Mutants Disrupt Glucocorticoid Receptor Ligand Binding and Destabilize Aporeceptor Complexes. Journal of Biological Chemistry. 270 (49): 29433-29438.

24. Bolanos-Garcia VM, Kiyomitsu T, D'Arcy S, Chirgadze DY, Grossmann JG, Matak-Vinkovic D, Venkitaraman AR, Yanagida M, Robinson CV, Blundell TL. 2009. The crystal structure of the N-terminal region of BUB1 provides insight into the mechanism of BUB1 recruitment to kinetochores. Structure. 17 (1): 105-16.

25. Borkovich KA, Farrelly FW, Finkelstein DB, Taulien J, Lindquist S. 1989. Hsp82 is an essential protein that is required in higher concentrations for growth of cells at higher temperatures. Mol Cell Biol. 9 (9): 3919–3930.

26. Boys BL, Konermann L. 2007. Folding and assembly of hemoglobin monitored by electrospray mass spectrometry using an on-line dialysis system. J Am Soc Mass Spectrom. 18 (1): 8-16.

27. Bricogne D. 1993. Direct phase determination by entropy maximization and likelihood ranking: status report and perspectives. Acta Crystallogr D Biol Crystallogr. 49 (Pt 1): 37-60.

28. Brown PH, Balbo A, Schuck P. 2009. On the analysis of sedimentation velocity in the study of protein complexes. Eur Biophys J. 38 (8): 1079-1099.

29. Buchan DW, Ward SM, Lobley AE, Nugent TC, Bryson K, Jones, DT. 2010. Protein annotation and modelling servers at University College London. Nucl Acids Res. 38 (Web Server issue): W563-W568.

30. Büttner Cr, Sorg I, Cornelis GR, Heinz DW, Niemann HH. 2008. Structure of the Yersinia enterocolitica type III secretion translocator chaperone SycD. J Mol Biol. 375 (4): 997-1012.

31. Catlett MG, Kaplan KB. 2006. Sgt1p is a unique co-chaperone that acts as a client adaptor to link Hsp90 to Skp1p. J Biol Chem. 281 (44): 33739-48.

32. Chadli A, Bouhouche I, Sullivan W, Stensgard B, McMahon N, Catelli MG, Toft DO. 2000. Dimerisation and N-terminal domain proximity underlie the function of the molecular chaperone heat shock protein 90. Proc Natl Acad Sci USA. 97 (23): 12524-12529.

33. Chen S, Sullivan WP, Toft DO, Smith DF. 1998. Differential Interactions of p23 and the TPR-containing proteins Hop, Cyp40, FKBP52 and FKBP51 with Hsp90 mutants. Cell Stress Chaperones. 3 (2): 118-129.

34. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 66 (Pt 1): 12-21.

35. Chen Y, Baker RE, Keith RC, Harris K, Stoler S, Fitzgerald-Hayes M. 2000. The N terminus of the centromere H3-like protein Cse4p performs an essential function distinct from that of the histone fold domain. Mol Cell Biol. 20 (18): 7037-7048.

36. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Karra K, Krieger CJ, Miyasato SR, Nash RS, Park J, Skrzypek MS, Simison M, Weng S, Wong ED. 2012. Saccharomyces Genome Database: the genomics resource of yeast. Nucl Acids Res. 40 (Database issue): D700-5.

37. Chinkers M. 2001. Protein phosphatase 5 in signal transduction. Trends Endocrinol Metab. 12 (1): 28-32.

38. Cho Uhn-Soo, Harrison SC. 2011. Ndc10 is a platform for inner kinetochore assembly in yeast. Nat Struct Mol Biol. 19 (1): 48-55.

39. Cliff MJ, Harris R, Barford D, Ladbury JE, Williams MA. 2006. Conformational diversity in the TPR-mediated interaction of protein phosphatase 5 with Hsp90. Structure. 14 (3): 415-426.

40. Cliff MJ, Williams MA, Brooke-Smith J, Barford D, Ladbury JE. 2005. Molecular recognition via coupled folding and binding in a TPR domain. J Mol Biol. 346 (3): 717-732.

41. Cohen RL, Espelin CW, De Wulf P, Sorger PK, Harrison SC, Simons KT. 2008. Structural and functional dissection of Mif2p, a conserved DNA-binding kinetochore protein. 19 (10): 4480-4491.

42. Cole JL, Lary JW, Moody T, Laue TM. 2008. Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. Methods Cell Biol. 84: 143-179.

43. Cowtan K. 1994. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography. 31: 34-38.

44. Cowtan K. 2000. General quadratic functions in real and reciprocal space and their application to likelihood phasing. Acta Crystallogr D Biol Crystallogr. 56 (Pt 12): 1612-1621.

45. Cowtan K. 2006. The *Buccaneer* software for automated model building. 1. Tracing protein chains. Acta Crystallogr D Biol Crystallogr. 62 (Pt 9): 1002-1011.

46. Crevel G, Bates H, Huikeshoven H, Cotterill S. 2001. The Drosophila Dpit47 protein is a nuclear Hsp90 co-chaperone that interacts with DNA polymerase alpha. J Cell Sci. 114 (Pt. 11): 2012-2025.

47. Cuff AL, Sillitoe I, Lewis T, Clegg AB, Rentzsch R, Furnham N, Pellegrini-Calace M, Jones D, Thornton J, Orengo CA. 2011. Extending CATH: increasing coverage of the protein structure universe and linking structure with function. Nucl Acids Res. 39 (Database issue): D420-6.

48. D'Andrea LD, Regan L. 2003. TPR proteins: the versatile helix. Trends in Biochem Sci. 28 (12): 655-62.

49. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB 3rd, Snoevink J, Richardson JS, Richardson DC. 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 35 (Web Server issue): W375-W383.

50. Davies AE, Kaplan KB. 2010. Hsp90-Sgt1 and Skp1 target human Mis12 complexes to ensure efficient formation of kinetochore-microtubule binding sites. J Cell Biol. 189 (2): 261-274.

51. Di Tullio A, Reale S, De Angelis F. 2005. Molecular recognition by mass spectrometry. J Mass Spectrom. 40 (7): 845-865.

52. Doheny KF, Sorger PK, Hyman AA, Tugendreich S, Spencer F, Hieter P. 1993. Identification of Essential Components of the S. cerevisiae Kinetochore. Cell. 73 (4): 761-774.

53. Dubacq C, Guerois R, Courbeyrette R, Kitagawa K, Mann C. 2002. Sgt1p Contributes to Cyclic AMP Pathway Activity and Physically Interacts with Adenylyl Cyclase Cyr1p/Cdc35p in Yeast. Eukaryot Cell. 1 (4): 568-582.

54. Eastburn DJ, Zegers MM, Mostov KE. 2012. Scrib regulates HGF-mediated epithelial morphogenesis and is stabilized by Sgt1-HSP90. J Cell Sci. *Epub ahead of print*.

55. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 66 (Pt. 4): 486-501.

56. Espelin CW, Kaplan KB, Sorger PK. 1997. Probing the architecture of a simple kinetochore using DNA-protein crosslinking. J Cell Biol. 139 (6): 1383-1396.

57. Flom GA, Langner E, Johnson JL. 2012. Identification of an Hsp90 mutation that selectively disrupts cAMP/PKA signaling in Saccharomyces cerevisiae. Curr Genet. 58 (3): 149-163.

58. Fromont-Racine M, Rain JC, Legrain P. 1997. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. Nat Genet. 16 (3): 277-282.

59. Funk M, Hegemann JH, Philippsen P. 1989. Chromatin digestion with restriction endonucleases reveals 150-160 bp of protected DNA in the centromere of chromosome XIV in *Saccharomyces cerevisiae*. Mol Gen Genet. 219 (1-2): 153-160.

60. Gardner RB, Poddar A, Yellman C, Tavormina PA, Monteagudo MC, Burke DJ. 2001. The spindle checkpoint of the yeast Saccharomyces cerevisiae requires kinetochore function and maps to the CBF3 domain. Genetics. 157 (4): 1493-1502.

61. Geller R, Vignuzzi M, Andino R, Frydman J. 2007. Evolutionary constraints on chaperone-mediated folding provide an antiviral approach refractory to development of drug resistance. Genes Dev. 21 (2): 195-205.

62. Geng F, Wenzel S, Tansey WP. 2012. Ubiquitin and Proteasomes in Transcription. Annu Rev Biochem. 81: 177-201.

63. Gonzalez JM, Velez M, Jimenez M, Alfonso C, Schuck P, Mingorance J, Vicente M, Minton AP, Rivas G. 2005. Cooperative behaviour of Escherichia coli cell-division protein FtsZ assembly involves the preferential cyclization of long single-stranded fibrils. Proc Natl Acad Sci USA. 102 (6): 1895-1900.

64. Gouet P, Courcelle E, Stuart DI, Metoz F. 1999. ESPript: multiple sequence alignments in PostScript. Bioinformatics. 15: 305-308.

65. Glowczewski L, Yang P, Kalashnikova T, Santisteban MS, Smith MM. 2000. Histone-histone interactions and centromere function. Mol Cell Biol. 20: 5700-5711.

66. Hammersley AP, Brown K, Burmeister W, Claustre L, Gonzalez A, McSweeney S, Mitchell E, Moy JP, Svensson SO, Thompson AW. 1997. Calibration and application of an X-ray image intensifier/charge-coupled device detector for monochromatic macromolecular crystallography. J Synchrotron Radiat. 4 (Pt. 2): 67-77.

67. Hanahan D. 1985. DNA Cloning, A Practical Approach. Volume 1. IRL Press, Ltd. London, UK. 109.

68. Hanahan D. 1983. Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol. 166: 557-577.

69. Hao B, Oehlmann S, Sowa ME, Harper JW, Pavletich NP. 2007. Structure of a Fbw7-Skp1-Cyclin E Complex: Multisite-Phosphorylated Substrate Recognition by SCF Ubiquitin Ligases. Mol Cell. 26 (1): 131-143.

70. Hao B, Zheng N, Schulmann BA, Wu G, Miller JJ, Pagano M, Pavletich NP. 2005. Structural basis for the Cks1-dependent recognition of p27(Kip1) by the SCF(Skp2) ubiquitin ligase. Mol Cell. 20: 9-19.

71. Harmand D. 2006. F-box proteins: more than bait for the SCF? Cell Division. 1: 30.

72. Harris SF, Shiau AK, Agard DA. 2004. The Crystal Structure of the Carboxy-Terminal Dimerisation Domain of htpG, the Escherichia coli HSP90, Reveals a Potential Substrate Binding Site. Structure. 12: 1087-1097.

73. Hartl FU, Bracher A, Hayer-Hartl M. 2011. Molecular chaperones in protein folding and proteostasis. Nature. 475: 324-332.

74. Hawle P, Siepmann M, Harst A, Siderius M, Reusch HP, Oberman WMJ. 2006. The Middle Domain of HSP90 Acts as a Discriminator between Different Types of Client Proteins. Mol Cell Biol. 26 (22): 8385-8395.

75. He X, Rines DR, Espelin CW, Sorger PK. 2001. Molecular Analysis of Kinetochore-Microtubule Attachment in Yeast. 106 (2): 195-206.

76. Hessling M, Richter K, Buchner J. 2009. Dissection of the ATP-induced conformational cycle of the molecular chaperone Hsp90. Struct Mol Biol. 16: 287-293.

77. Hieter P, Connelly C. 1996. Yeast *SKP1* encodes an evolutionarily conserved kinetochore protein required for cell cycle progression. Cell. 86: 275-285.

78. Höhfeld J, Minami Y, Hartl FU. 1995. Hip, a novel cochaperone involved in the eukaryotic Hsc70/Hsp40 reaction cycle. Cell. 83 (4): 589-598.

79. Hoffmann DB, Pearson CG, Yen TJ, Howell BJ, Salmon ED. 2001. Microtubule-dependent Changes in Assembly of Microtubule Motor Proteins and Mitotic Spindle Checkpoint Proteins at PtK1 Kinetochores. Mol Biol Cell. 12 (7): 1995-2009.

80. Holm L, Park J. 2000. DaliLite workbench for protein structure comparison. Bioinformatics. 16 (6): 566-567.

81. Hori T, Amano M, Suzuki A, Backer CB, Welburn JP, Dong Y, McEwen BF, Shang W-H, Suzuki E, Okawa K, Cheeseman IM, Fukagawa T. 2008. CCAN Makes Multiple Contacts with Centromeric DNA to Provide Distinct Pathways to the Outer Kinetochore. Cell. 135 (6): 1039-1052.

82. Howlett GJ, Minton AP, Rivas G. 2006. Analytical ultracentrifugation for the study of protein association and assembly. Current Opinion in Chemical Biology. 10: 430-436.

83. Huang DT, Paydar A, Zhuang M, Waddell MB, Holton JM, Schulman BA. 2005. Structural basis for recruitment of Ubc12 by an E2 binding domain in NEDD8's E1. Mol Cell. 17 (3): 341-350.

84. Immormino RM, Dollins DE, Shaffer PI, Soldano KL, Walker MA, Gewirth DT. 2004. Ligand-induced conformational shift in the N-terminal domain of GRP94, an Hsp90 chaperone. J Biol Chem. 279: 46162-46171.

85. Incardona MF, Bourenkov GP, Levik K, Pieritz RA, Popov AN, Svensson O. EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. J Synchrotron Rad. 16: 872-879.

86. Jackson SE. 2012. Hsp90: Structure and Function. Top Curr Chem. *Epub ahead of print*.

87. Jehn B, Niedenthal R, Hegemann JH. 1991. In vivo analysis of the Saccharomyces cerevisiae centromere CDEIII sequence: requirements for mitotic chromosome segregation. Mol Cell Biol. 11: 5212-5221.

88. Jiang W, Lechner J, Carbon J. 1993. Isolation and characterisation of a gene (CBF2) specificying a protein component of the yeast kinetochore. J Cell Biol. 121 (3): 513-519.

89. Johnson BD, Schumacher RJ, Ross ED, Toft DO. 1998. Hop modulates Hsp70/Hsp90 interactions in protein folding. J Biol Chem. 273 (6): 3679-3686.

90. Johnson LN, Lewis RJ. 2001. Structural Basis for Control by Phosphorylation. Chem Rev. 101: 2209-2242.

91. Jonkers W, Rep M. 2009. Lessons from Fungal F-Box Proteins. Eukaryot Cell. 8 (5): 677-695.

92. Kantardjieff KA, Rupp B. 2003. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. Protein Sci. 12 (9): 1865-1871.

93. Kabsch W. 2009. XDS. Acta Crystallogr D Biol Crystallogr. D66: 125-132.

94. Kadota Y, Amigues B, Ducassou L, Madoui H, Ochsenbein F, Guerois R, Shirasu K. 2008. Structural and functional analysis of SGT1-HSP90 core complex required for innate immunity in plants. EMBO Rep. Available online: 03 October 2008.

95. Kadota Y, Shirasu K, Guerois R. 2010. NLR sensors meet at the SGT1-HSP90 crossroad. Trends Biochem Sci. 35 (4): 199-207.

96. Kaplan KB, Hyman AA, Sorger PK. 1997. Regulating the yeast kinetochore by ubiquitin-dependent degradation and Skp1p-mediated phosphorylation. Cell. 91: 491-500.

97. Keating P, Rachidi N, Tanaka TU, Stark MJ. 2009. Ipl1-dependent phosphorylation of Dam1 is reduced by tension applied on kinetochores. J Cell Sci. 122 (Pt 23): 4375-4382.

98. Keegan RM, Winn MD. 2007. Automated search-model discovery and preparation for structure solution by molecular replacement. Acta Crystallogr D Biol Crystallogr. 63: 447-457.

99. Kiefer F, Arnold K, Bordoli L, Schwede T. 2009. The SWISS-MODEL Repository and Associated Resources. Nucl Acids Res. 37: D387-D392.

100. Kim K, Oh J, Han D, Kim EE, Lee B, Kim Y. 2006. Crystal structure of PilF: functional implication in the type 4 pilus biogenesis in Pseudomonas aeruginosa. Biochem Biophys Res Commun. 340 (4): 1028-1038.

101. Kipreos ET, Pagano M. 2000. The F-box protein family. Gen Biol. 1 (5): 1-7.

102. Kitagawa K, Abdulle R, Bansal PK, Cagney G, Fields S, Hieter P. 2003. Requirement of Skp1-Bub1 Interaction for Kinetochore-Mediated Activation of the Spindle Checkpoint. Mol Cell. 11: 1201-1213.

103. Kitagawa K, Skowyra D, Elledge SJ, Harper JW, Hieter P. 1999. SGT1 Encodes an Essential Component of the Yeast Kinetochore Assembly Pathway and a Novel Subunit of the SCF Ubiquitin Ligase Complex. Mol Cell. 4: 21-33.

104. Kiyomitsu T, Obuse C, Yanagida M. 2007. Human Blinkin/AF15q14 Is Required for Chromosome Alignment and the Mitotic Checkpoint through Direct Interaction with Bub1 and BubR1. Dev Cell. 13 (5): 663-676.

105. Kobe B, Kajava AV. 2001. The leucine-rich repeat as a protein recognition motif. Curr Opin Struct Biol. 11 (6): 725-732.

106. Kostanski LK, Keller DM, Hamielec AE. 2004. Size-exclusion chromatography – a review of calibration technologies. J Biochem Biophys Methods. 58: 159-186.

107. Koulov AV, LaPointe P, Lu B, Razvi A, Coppinger J, Dong M, Matteson J, Laister R, Arrowsmith C, Yates JR, Balch WE. 2010. Biological and Structural Basis for Aha1 Regulation of Hsp90 ATPase Activity in Maintaining Proteostasis in the Human Disease Cystic Fibrosis. Mol Biol Cell. 21 (6): 871-884.

108. Krachler AM, Sharma A, Cauldwell A, Papadakos G, Kleanthous C. 2010. TolA modulates the oligeromic status of YbgF in the bacterial periplasm. J Mol Biol. 403 (2): 270-285.

109. Krachler AM, Sharma A, Kleanthous C. 2010. Self-association of TPR domains: Lessons learned from a designed, consensus-based TPR oligomer. Proteins. 78 (9): 2131-43.

110. Krissinel E, Henrick K. 2007. Inference of macromolecular assemblies from crystalline state. J Mol Biol. 372: 774-797.

111. Krukenberg KA, Street TO, Lavery LA, Agard DA. 2011. Conformational dynamics of the molecular chaperone Hsp90. Q Rev Biophys. 44 (2): 229-255.

112. Kundrat L, Regan L. 2010. Balance Between Folding and Degradation for Hsp90-Dependent Client Proteins: A Key Role for CHIP. Biochemistry. 49 (35): 7428-7438.

113. Larson JL, Ko E, Miranker AD. 2000. Direct measurement of islet amyloid polypeptide fibrillogenesis by mass spectrometry. Protein Science. 9: 427-431.

114. Laskowski RA. 2009. PDBsum new things. Nucl Acids Res. 37: D355-D359.

115. Laue TM, Shah BD, Ridgeway TM, Pelletier SL. 1992. Analytical Ultracentrifugation in Biochemistry and Polymer Science. Royal Society of Chemistry.

116. Lebowitz J, Lewis MS, Schuck P. 2002. Modern analytical ultracentrifugation in protein science: A tutorial review. Protein Science. 11: 2067-2079.

117. Lechner J. 1994. A zinc finger protein, essential for chromosome segregation, constitutes a putative DNA binding subunit of the *Saccharomyces cerevisiae* kinetochore complex, Cbf3. EMBO J. 13 (21): 5203-5211.

118. Lee CT, Graf C, Mayer FJ, Richter SM, Mayer MP. 2012. Dynamics of the regulation of Hsp90 by the co-chaperone Sti1. EMBO J. 31 (6): 1518-1528.

119. Lee I, Schindelin H. 2008. Structural insights into E1-catalyzed ubiquitin activation and transfer to conjugating enzymes. Cell. 134 (2): 268-278.

120. Lee YT, Jacob J, Michowski W, Nowotny M, Kuznicki J, Chazin WJ. 2004. Human Sgt1 binds Hsp90 through the CHORD-Sgt1 domain and not the tetratricopeptide repeat domain. J Biol Chem. 279 (16): 16511-16517.

121. Lingelbach LB, Kaplan KB. 2004. The Interaction between Sgt1p and Skp1p is Regulated by HSP90 Chaperones and Is Required for Proper CBF3 Assembly. Mol Cell Biol. 24(20): 8938-8950.

122. Liu XS, Song B, Tang J, Liu W, Kuang S, Liu X. 2012. Plk1 phosphorylates Sgt1 at the kinetochores to promote the timely kinetochore-microtubule attachment. Mol Cell Biol. *Epub ahead of print*.

123. Lokareddy RK, Lunelli M, Eilers B, Wolter V, Kolbe M. 2010. Combination of Two Separate Binding Domains Defines Stoichiometry between Type III Secretion System Chaperone IpgC and Translocator Protein IpaB. J Biol Chem. 285: 39965-39975.

124. Lotz GP, Lin H, Harst A, Obermann WM. 2003. Aha1 Binds to the Middle Domain of Hsp90, Contributes to Client Protein Activation, and Stimulates the ATPase Activity of the Molecular Chaperone. J Biol Chem. 278: 17228-17235.

125. MacGurn JA, Hsu PC, Emr SD. 2012. Ubiquitin and membrane protein turnover: from cradle to grave. Annu Rev Biochem. 81: 231-259.

126. Main ER, Xiong Y, Cocco MJ, D'Andrea L, Regan L. 2003. Design of stable alpha-helical arrays from an idealized TPR motif. Structure. 11 (5): 497-508.

127. Marmorstein R, Carey M, Ptashne M, Harrison S. 1992. DNA recognition by GAL4: structure of a protein-DNA complex. Nature. 356: 408-414.

128. Mayer CL, Snyder WK, Swietlicka MA, VanSchoiack AD, Austin CR, McFarland BJ. 2009. Size-exclusion chromatography can identify faster-associating protein complexes and evaluate design strategies. BMC Research Notes. 2: 135.

129. Mayer MP, Bukau B. 2005. Hsp70 chaperones: Cellular functions and molecular mechanism. Cell Mol Life Sci. 62: 670-684.

130. McAinsh AD, Meraldi P. 2011. The CCAN Complex: Linking centromere specification to control of kinetochore-microtubule dynamics. Semin Cell Dev Biol. 22 (9): 946-952.

131. McLaughlin SH, Smith HW, Jackson SE. 2002. Stimulation of the weak ATPase activity of human Hsp90 by a client protein. J Mol Biol. 315: 787-798.

132. McLaughlin SH, Sobott F, Yao ZP, Zhang W, Nielsen PR, Grossman JG, Laue ED, Robinson CV, Jackson SE. 2006. The co-chaperone p23 arrests the Hsp90 ATPase cycle to trap client proteins. J Mol Biol. 356: 746-758.

133. McLaughlin SH, Ventouras L, Lobbezoo B, Jackson SE. 2004. Independent ATPase Activity of Hsp90 Subunits Creates a Flexible Assembly Platform. J Mol Biol. 344: 813-826.

134. McLeland ML, Gardner RD, Kallio MJ, Daum JR, Gorbsky GJ, Burke DJ, Stukenberg PT. 2003. The highly conserved Ndc80 complex is required for kinetochore assembly, chromosome segregation, and spindle checkpoint activity. Genes Dev. 17: 101-114.

135. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. *Phaser* crystallographic software. J Appl Crystallogr. 40: 658-674.

136. Meimaridou E, Gooljar SB, Chapple JP. 2009. From hatching to dispatching: the multiple cellular roles of the Hsp70 molecular chaperone machinery. J Mol Endocrinol. 42 (1): 1-9.

137. Meluh PM, Yang P, Glowczewski L, Koshland D, Smith MM. 1998. Cse4p Is a Component of the Core Centromere of *Saccharomyces cerevisiae*. Cell. 94 (5): 607-613.

138. Meraldi P, McAinsh AD, Rheinbay E, Sorger PK. 2006. Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. Genome Biol. 7 (3): R23.

139. Meyer P, Prodromou C, Chunyan L, Hu B, Roe SM, Vaughan CK,Vlasic I, Panaretou B, Piper PW, Pearl LH. 2004. Structural basis for the recruitment of the ATPase activator Aha1 to the Hsp90 chaperone machinery. EMJO J. 23: 1402-1410.

140. Meyer P, Prodromou C, Hu B, Vaughan C, Roe SM, Panaretou B, Piper PW, Pearl LH. 2003. Structural and functional analysis of the middle segment of hsp90: implications for ATP hydrolysis and client protein and cochaperone interactions. Mol Cell. 11 (3): 647-658.

141. Meyer P, Prodromou C, Liao C, Hu B, Roe SM, Vaughan CK, Vlasic I, Panaretou B, Piper PW, Pearl LH. 2004. Structural basis for the recruitment of the ATPase activator Aha1 to the Hsp90 chaperone machinery. EMBO Journal. 23(6): 1402-1410.

142. Mickler M, Hessling M, Ratzke C, Buchner J, Hugel T. 2009. The large conformational changes of Hsp90 are only weakly coupled to ATP hydrolysis. Nat Struct Mol Biol. 16 (3): 281-286.

143. Minton AP. 2000. Quantitative characterisation of reversible macromolecular associations via sedimentation equilibrium: an introduction. Experimental and Molecular Medicine. 32(1): 1-5.

144. Miyata Y. 2009. Protein kinase CK2 in health and disease: CK2: the kinase controlling the Hsp90 chaperone machinery. Cell Mol Life Sci. 66 (11-12): 1840-1849.

145. Morra G, Verkhivker G, Colombo G. 2009. Modeling Signal Propagation Mechanisms and Ligand-Based Conformational Dynamics of the Hsp90 Molecular Chaperone Full-Length Dimer. PLoS Comput Biol. 5 (3): e1000323.

146. Morris RJ, Zwart PH, Cohen S, Fernandez FJ, Kakaris M, Kirilova O, Vonrhein C, Perrakis A, Lamzin VS. 2004. Breaking good resolutions with ARP/wARP. J Synchrotron Radiat. 11 (Pt 1): 56-59.

147. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin A. 2011. *REFMAC5* for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr. D67: 355-367.

148. Navaza J. 1994. AMoRe: an automated package for molecular replacement. Acta Crystallogr A Biol Crystallogr. 50: 157-163.

149. Neckers L, Workman P. 2012. Hsp90 molecular chaperone inhibitors: are we there yet? Clin Cancer Res. 18 (1): 64-76.

150. Niikura Y, Ohta S, Vandenbeldt KJ, Abdulle R, McEwen BF, Kitagawa K. 2006. 17-AAG, an Hsp90 inhibitor, causes kinetochore defects: a novel mechanism by which 17-AAG inhibits cell proliferation. Oncogene. 25: 4133-4146.

151. Norman AW, Mizwicki MT, Norman DP. 2004. Steroid-hormone rapid actions, membrane receptors and a conformational ensemble model. Nat Rev Drug Discov. 3 (1): 27-41.

152. Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol. 302 (1): 205-217.

153. Nyarko A, Mosbahi K, Rowe AJ, Leech A, Botër M, Shirasu K, Kleanthous C. 2007. TPR-Mediated Self-Association of Plant SGT1. Biochemistry. 46: 11331-11341.

154. Oliva A, Llabrés M, Farina JB. 2001. Comparative study of protein molecular weights by size-exclusion chromatography and laser-light scattering. J Pharm Biomed Anal. 25 (5-6): 833-841.

155. Olsen SK, Capili AD, Lu X, Tan DS, Lima CD. 2009. Active site remodelling accompanies thioester bond formation in the SUMO E1. Nature. 463: 906-912.

156. Onuoha SC, Coulstock Et, Grossmann JG, Jackson SE. 2008. Structural Studies on the Co-chaperone Hop and Its Complexes with Hsp90. J Mol Biol. 379: 732-744.

157. Orlicky S, Tang X, Neduva V, Elowe N, Brown ED, Sicheri F, Tyers M. 2010. An allosteric inhibitor of substrate recognition by SCF(cdc4) ubiquitin ligase. Nat Biotechnol. 28 (7): 733-737.

158. Ortenga A, Amoros D, Garcia de la Torre J. 2011. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. Biophys J. 101 (4): 892-898.

159. Ortiz J, Stemmann O, Rank S, Lechner J. 1999. A putative protein complex consisting of Ctf19, Mcm21, and Okp1 represents a missing link in the yeast kinetochore. Genes Dev. 13 (9): 1140-1155.

160. Panjikar S, Parthasarathy V, Lamzin VS, Weiss MS, Tucker PA. 2005. *Auto-Rickshaw* – An automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. Acta Cryst. D61: 449-457.

161. Panjikar S, Parthasarathy V, Lamzin VS, Weiss MS, Tucker PA. 2009. On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination. Acta Crystallogr D Biol Crystallogr. 65: 1089-1097.

162. Pearl LH. 2005. Hsp90 and Cdc37 – a chaperone cancer conspiracy. Curr Opin Genet Dev. 15 (1): 55-61.

163. Pearl LH, Prodromou C. 2006. Structure and Mechanism of the Hsp90 Molecular Chaperone Machinery. Annual Reviews of Biochemistry. 75: 271-294.

164. Peitsch MC. 1995. Protein modelling by E-mail. Bio Tech. 13: 658-660.

165. Pekkala M, Hieta R, Bergmann U, Kivirikko KI, Wierenga RK, Myllyharju J. 2004. The peptide-substrate-binding domain of collagen prolyl 4-hydroxylases is a tetratricopeptide repeat domain with functional aromatic residues. J Biol Chem. 279: 52255-52261.

166. Perriches T, Single MR. 2012. Structure of Yeast Kinetochore Ndc10 DNA-binding Domain Reveals Unexpected Evolutionary Relationship to Tyrosine Recombinases. J Biol Chem. 287 (7): 5173-5179.

167. Pietrasanta L, Thrower D, Hsieh W, Rao S, Stemmann O, Lechner J, Carbon J, Hansma H. 1999. Probing the Saccharomyces cerevisiae centromeric DNA (CEN DNA)-binding factor 3 (CBF3) kinetochore complex by atomic force microscopy. PNAS. 96: 3757-3762.

168. Pinna LA. 1990. Casein kinase 2: an 'eminence grise' in cellular regulation? Biochim Biophys Acta. 1054 (3): 267-284.

169. Prodromou C, Panaretou B, Chohan S, Siligardi G, O'Brien R, Ladbury JE, Roe SM, Piper PW, Pearl LH. 2000. The ATPase cycle of Hsp90 drives a molecular 'clamp' via transient dimerisation of the N-terminal domains. EMBO J. 19 (16): 4383-4392.

170. Prodromou C, Roe SM, O'Brien R, Ladbury JE, Piper PW, Pearl LH. 1997. Identification and Structural Characterisation of the ATP/ADP-Binding Site in the Hsp90 Molecular Chaperone. Cell. 90: 65-75.

171. Prodromou C, Piper PW, Pearl LH. 1996. Expression and Crystallization of the Yeast Hsp82 Chaperone, and the Preliminary X-Ray Diffraction Studies of the Amino-Terminal Domain. Proteins. 25: 517-522.

172. Prodromou C, Roe SM, Piper PW, Pearl LH. 1997. A molecular clamp in the crystal structure of the N-terminal domain of the yeast Hsp90 chaperone. Nature. 4(6): 477-482.

173. Prodromou C, Siligardi G, O'Brien R, Woolfson DN, Regan L, Panaretou B, Ladbury JE, Piper PW, Pearl LH. 1999. Regulation of Hsp90 ATPase activity by tetratricopeptide repeat (TPR)-domain co-chaperones. EMBO Journal. 18(3): 754-762.

174. Punta M, Coggill PC, Eberhardt PY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD. 2012. The Pfam protein families database. Nucleic Acids Res. 40 (Database Issue): D290-D301.

175. Purvis A, Singleton MR. 2008. Insights into kinetochore-DNA interactions from the structure of Cep3Δ. EMBO Reports. 9 (1): 56-62.

176. Ratzke C, Berkemeier F, Hugel T. 2012. Heat shock protein 90's mechanochemical cycle is dominated by thermal fluctuations. Proc Natl Acad Sci USA. 109 (1): 161-166.

177. Retzlaff M, Hagn F, Mitschke L, Hessling M, gugel F, Kessler H, Richter K, Buchner J. 2010. Asymmetric Activation of the Hsp90 Dimer by Its Cochaperone Aha1. Mol Cell. 37: 344-354.

178. Richter K, Moser S, Hagn F, Friedrich R, Hainzl O, Heller M, Schlee S, Kessler H, Reinstein J, Buchner J. 2006. Instrinsic Inhibition of the Hsp90 ATPase Activity. J Biol Chem. 281: 11301-11311.

179. Richter K, Muschler P, Hainzl O, Buchner J. 2001. Coordinated ATP hydrolysis by the Hsp90 dimer. J Biol Chem. 276 (36): 33689-33696.

180. Richter K, Muschler P, Hainzl O, Reinstein J, Buchner J. 2003. Sti1 is a non-competitive inhibitor of the Hsp90 ATPase. Binding prevents the N-terminal dimerisation reaction during the ATPase cycle. Journal of Biological Chemistry. 278(12): 10328-10333.

181. Rodrigo-Brenni MC, Thomas S, Bouck DC, Kaplan KB. 2004. Sgt1p and Skp1p Modulate the Assembly and Turnover of CBF3 Complexes Required for Proper Kinetochore Function. Molecular Biology of the Cell. 15: 3366-3378.

182. Roe SM, Ali MM, Meyer P, Vaughan CK, Panaretou B, Piper PW, Prodromou C, Pearl LH. 2004. The Mechanism of Hsp90 regulation by the protein kinase-specific cochaperone p50(cdc37). Cell. 116 (1): 87-98.

183. Russell ID, Grancell AS, Sorger PK. 1999. The Unstable F-Box Protein p58-Ctf13 Forms the Structural Core of the CBF3 Kinetochore Complex. J Cell Biol. 145 (5): 933-950.

184. Rutherford SL, Lindquist S. 1998. Hsp90 as a capacitor for morphological evolution. 396 (6709): 336-342.

185. Sato S, Fujita N, Tsuruo T. 2000. Modulation of Akt kinase activity by binding to Hsp90. Proc Natl Acad Sci USA. 97 (20): 10832-10837.

186. Scheibel T, Siegmund HI, Jaenicke R, Ganz P, Lilie H, Buchner J. 1999. The charged region of Hsp90 modulates the function of the N-terminal domain. PNAS. 96(4): 1297-1302.

187. Scheufler C, Brinkler A, Bourenkov G, Pegoraro S, Moroder L, Bartunik H, Hartl FU, Moarefi I. 2000. Structure of TPR Domain-Peptide Complexes: Critical Elements in the Assembly of the Hsp70-Hsp90 Multichaperone Machine. Cell. 101: 199-210.

188. Schreiner M, Niemann HH. 2012. Crystal structure of the Yersinia enterocolitica type III secretion chaperone SycD in complex with a peptide of the minor translocator YopD. BMC Struct Biol. 12 (1): 13.

189. Schuck P. 2000. Size distribution analysis of macromolecules by sedimentation velocity analytical ultracentrifugation and Lamm equation modeling. Biophys J. 78: 1606-1619.

190. Schulman BA, Carrano AC, Jeffrey PD, Bowen Z, Kinnucan ERE, Finnin MS, Elledge SJ, Harper JW, Pagano M, Pavletich NP. 2000. Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. Nature. 408: 381-386.

191. Schulze-Lefert P. 2004. Plant immunity: the origami of receptor activation. Curr Biol. 14: R22-R24.

192. Sharon M, Robinson CV. 2007. The Role of Mass Spectrometry in Structure Elucidation of Dynamic Protein Complexes. Annu Rev Biochem. 76: 167-193.

193. Sheldrick GM. 2010. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Crystallogr D Biol Crystallogr. 66: 479-485.

194. Shiau AK, Harris SF, Southworth DR, Agard DA. 2006. Structural Analysis of E. coli Hsp90 Reveals Dramatic Nucleotide-Dependent Conformational Arrangements. Cell. 127: 329-340.

195. Shirasu K. 2009. The HSP90-SGT1 Chaperone Complex for NLR Immune Sensors. Annual Reviews in Plant Biology. 60: 139-164.

196. Siergiejuk E, Scott DC, Schulman BA, Hofmann K, Kurz T, Peter M. 2009. Cullin neddylation and substrate-adaptors counteract SCF inhibition by the CAND1-like protein Lag2 in *Sacchromyces cerevisiae*. EMBO J. 28 (24): 3845-3856.

197. Singh M, Gonzales FA, Cascio D, Heckmann N, Chanfreau G, Feigon J. 2009. Structure and Functional Studies of the CS Domain of the Essential H/ACA Ribonucleotideparticle Assembly Protein SHQ1. Journal of Biological Chemistry. 284 (3): 1906-1916.

198. Sironi L, Mapelli M, Knapp S, De Antoni A, Jeang K-T, Musacchio A. 2002. Crystal structure of the tetrameric Mad1-Mad2 core complex: implications of a 'safety belt' binding mechanism for the spindle checkpoint. EMBO J. 12 (10): 2496-2506.

199. Smith MM. 2002. Centromeres and variant histones: what, where, when and why? Current Opinion in Cell Biology. 14: 279-285.

200. Sobott F, McCammon MG, Robinson CV. 2003. Gas-phase dissociation pathways of a tetrameric protein complex. Int J Mass Spectrom. 230: 193-200.

201. Southworth DR, Agard DA. 2008. Species-Dependent Ensembles of Conserved Conformational States Define the Hsp90 Chaperone ATPase Cycle. Molecular Cell. 32: 632-640.

202. Spiechowicz M, Zylicz A, Bieganowski P, Kuznicki J, Filipek A. 2007. Hsp70 is a new target of Sgt1 – an interaction modulated by S100A6. Biochem Biophys Res Commun. 357 (4): 1148-1153.

203. Spoel SH, Loake GJ. 2011. Redox-based protein modifications: the missing link in plant immune signalling. Curr Opin Plant Biol. 14 (4): 358-364.

204. Stebbins CE, Russo AA, Schneider C, Rosen N, Hartl FU, Pavletich NP. 1997. Crystal Structure of an Hsp90-Geldanamycin Complex: Targeting of a Protein Chaperone by an Antitumour Agent. Cell. 89: 239-250.

205. Stemmann O, Lechner J. 1996. The Saccharomyces cerevisiae kinetochore contains a cyclin-CDK complexing homologue, as identified by in vitro reconstitution. EMBO J. 15 (14): 3611-3620.

206. Stemmann O, Neidig A, Köcher T, Wilm M, Lechner J. 2002. Hsp90 enables Ctf13p/Skp1p to nucleate the yeast kinetochore. Proc Natl Acad Sci USA. 99 (13): 8585-8590.

207. Steensgaard P, Garrè M, Muradore I, Transidico P, Trigg EA, Kitagawa K, Earnshaw WC, Faretta M, Musacchio A. 2004. Sgt1 is required for human kinetochore assembly. EMBO Rep. 5 (6): 626-631.

208. Stogios PJ, Downs GS, Jauhal JJ, Nandra SK, Prive GG. 2005. Sequence and structural analysis of BTB domain proteins. Gen Biol. 6 (10): R82.1-R82.18.

209. Stoyan T, Carbon J. 2004. Inner Kinetochore of the Pathogenic Yeast *Candida glabrata*. Eukaryot Cell. 3 (5): 1154-1163.

210. Strunnikov AV, Kingsbury J, Koshland D. 1995. CEP3 Encodes a Centromere Protein of Saccharomyces cerevisiae. J Cell Biol. 128 (5): 749-760.

211. Suzuki M, Neutzner A, Tjandra N, Youle RJ. 2005. Novel structure of the N terminus in yeast Fis1 correlates with a specialized function in mitochondrial fission. J Biol Chem. 280 (22): 21444-52.

212. Tachiwana H, Kagawa W, Shiga T, Osakabe A, Miya Y, Saito K, Hayashi-Tanaka T, Oda T, Sato M, Park S, Kimura H, Kurumizaka H. 2011. Crystal structure of the human centromeric nucleosome containing CENP-A. Nature. 476: 232-235.

213. Tanaka TU, Rachidi N, Janke C, Pereira G, Galova M, Schiebel E, Stark MJ, Nasmyth K. 2002. Evidence that the Ipl1-Sli15 (Aurora kinase-INCENP) complex promotes chromosome bi-orientation by altering kinetochore-spindle pole connections. Cell. 108 (3): 317-329.

214. Takahashi A, Casais C, Ichimura K, Shirasu K. 2003. HSP90 Interacts with RAR1 and SGT1 and is Essential for RPS2-mediated Disease Resistance in Arabidopsis. PNAS. 100(20): 11777-11782.

215. Taylor P, Dornan J, Carrello A, Minchin RF, Ratajczak T, Walkinshaw MD. 2001. Two Structures of Cyclophilin 40: Folding and Fidelity in the TPR Domains. Structure. 9 (5): 431-438.

216. Terwilliger TC. 2000. Maximum likelihood density modification. Acta Crystallogr D Biol Crystallogr. 56: 965-972.

217. Turner GC, Varshavsky A. 2000. Detecting and measuring cotranslational protein degradation in vivo. Science. 289: 2117-2120.

218. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM. 2000. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature. 403(6770): 623-627.

219. Vagin A, Teplyakov A. 1997. MOLREP: an Automated Program for Molecular Replacement. J Appl Cryst. 30: 1022-1025.

220. Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J. 1993. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. J Mol Biol. 229 (1): 105-124.

221. Vaughan CK, Gohike U, Sobolt F, Good VM, Ali MM, Prodromou C, Robinson CV, Saibil HR, Pearl LH. 2006. Structure of an Hsp90-Cdc37-Cdk4 complex. Mol Cell. 23 (5): 697-707.

222. Vaughan CK, Piper PW, Pearl LH, Prodromou C. 2009. A common conformationally coupled ATPase mechanism for yeast and human cytoplasmic HSP90s. FEBS J. 276 (1): 199-209.

223. Vaughan CK, Mollapour M, Smith JR, Truman A, Hu B, Good VM, Panaretou B, Neckers L, Clarke PA, Workman P, Piper PW, Prodromou C, Pearl LH. 2008. Hsp90-dependent activation of protein kinases is regulated by chaperone-targeted dephosphorylation of Cdc37. Mol Cell. 31 (6): 886-895.

224. Vainshtein BK. 2003. Modern Crystallography 1: Fundamentals of Crystals, Symmetry, and Methods of Structural Crystallography. Second Edition. Springer.

225. Walhout AJ, Reboul J, Shtanko O, Bertin N, Vaglio P, Ge H, Lee H, Doucette-Stamm L, Gunsalus KC, Schetter AJ, Morton DG, Kemphues KJ, Reinke V, Kim SK, Piano F, Vidal M. 2002. Integrating interacome, phenome, and transcriptome mapping data for the C. elegans germline. Curr Biol. 12(22): 1952-1958.

226. Welburn JPI, Cheeseman IM. 2008. Toward a Molecular Structure of the Eukaryotic Kinetochore. Developmental Cell. 15: 645-655.

227. Wenzel DM, Stoll KE, Klevit RE. 2010. E2s: Structurally Economical and Functionally Replete. Biochem J. 433 (1): 31-42.

228. Westermann S, Drubin DG, Barnes G. 2007. Structures and Functions of Yeast Kinetochore Complexes. Annu Rev Biochem. 76: 563-591.

229. Westermann S, Wang HW, Avila-Sakar A, Drubin DG, Nogales E, Barnes G. 2006. The Dam1 Kinetochore Ring Complex Moves Processively On Depolymerizing Microtubule Ends. Nature. 440(7083): 565-569.

230. Whitesell L, Lindquist S. 2009. Inhibiting the transcription factor HSF1 as an anticancer strategy. Expert Opin Ther Targets. 13 (4): 469-478.

231. Wigley DB, Davies GJ, Dodson EJ, Maxwell A, Dodson G. 1991. Crystal Structure of an N-terminal Fragment of the DNA Gyrase B protein. Nature. 351: 624-629.

232. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF. 1999. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol. 112: 531-552.

233. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS,Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. 2011. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr. 67 (Pt. 4): 235-242.

234. Winter G. 2010. xia2: an expert system for macromolecular crystallography data reduction. J Appl Cryst. 43: 186-190.

235. Wu H, Yang WP, Barbas CF 3rd. 1995. Building zinc fingers by selection: toward a therapeutic application. Proc Natl Acad Sci USA. 92 (2): 344-348.

236. Wu J, Luo S, Jiang H, Li H. 2005. Mammalian CHORD-containing protein 1 is a novel heat shock protein 90-interacting protein. FEBS Letters. 579: 421-426.

237. Wu Y, McCombs D, Nagy L, DeLucas L, Sha B. 2006. Preliminary X-ray crystallographic studies of yeast mitochondrial protein Tom70p. Acta Crystallogr Sect F Struct Biol Cryst Commun. 62 (Pt 3): 265-267.

238. Xu W, Marcu M, Yuan X, Mimnaugh E, Patterson C, Neckers L. 2002. Chaperone-dependent E3 ubiquitin ligase CHIP mediates a degradative pathway for c-ErbB2/Neu. Proc Natl Acad Sci USA. 99 (20): 12847-12852.

239. Yang J, Roe SM, Cliff MJ, Williams MA, Ladbury JE, Cohen PT, Barford D. 2005. Molecular basis for TPR domain-mediated regulation of protein phosphatase 5. EMBO J. 24: 1-10.

240. Yoshida Y, Murakami A, Tanaka K. 2011. Skp1 stabilises the conformation of F-box proteins. Biochem Biophys Res Commun. 410 (1): 24-28.

241. Young JC, Obermann WM, Hartl FU. 1998. Specific binding of tetratricopeptide repeat proteins to the C-terminal 12-kDa domain of hsp90. J Biol Chem. 273 (29): 18007-18010.

242. Young JC, Schneider C, Hartl FU. 1997. In vitro evidence that hsp90 contains two independent chaperone sites. FEBS Letters. 418 (1-2): 139-143.

243. Zabka M, Lesniak W, Prus W, Kuznicki J, Filipek A. 2008. Sgt1 has co-chaperone properties and is up-regulated by heat shock. Biochemical and Biophysical Research Communications. 370: 179-183.

244. Zanata SM, Lopes MH, Mercandante AF, Hajj GN, Chiarini LB, Nomizo R, Freitas AR, Cabral AL, Lee KS, Juliano MA, de Oliveira E, Jachieri SG, Burlingame A, Huang L, Linden R, Brentani RR, Martins VR. 2002. Stress-inducible protein 1 is a cell surface ligand for cellular prion that triggers neuroprotection. EMBO J. 21 (13): 3307-3316.

245. Zeytuni N, Zarivach R. 2012. Structural and Functional Discussion of the Tetra-Trico-Peptide Repeat, a Protein Interaction Module. Structure. 20 (3): 397-405.

246. Zhang KY, Cowtan K, Main P. 1997. Combining constraints for electron-density modification. Methods Enzymol. 277: 53-64.

247. Zhang M, Botër M, Li K, Kadota Y, Panaretou B, Prodroumou C, Shirasu K, Pearl LH. 2008. Structural and functional coupling of Hsp90- and Sgt1-centred multi-protein complexes. EMBO J. 27 (20): 2789-2898.

248. Zhang M, Kadota Y, Prodromou C, Shirasu K, Pearl LH. 2010. Structural Basis for Assembly of Hsp90-Sgt1-CHORD Protein Complexes: Implications for Chaperoning of NLR Innate Immunity Receptors. Mol Cell. 39 (2): 269-281.

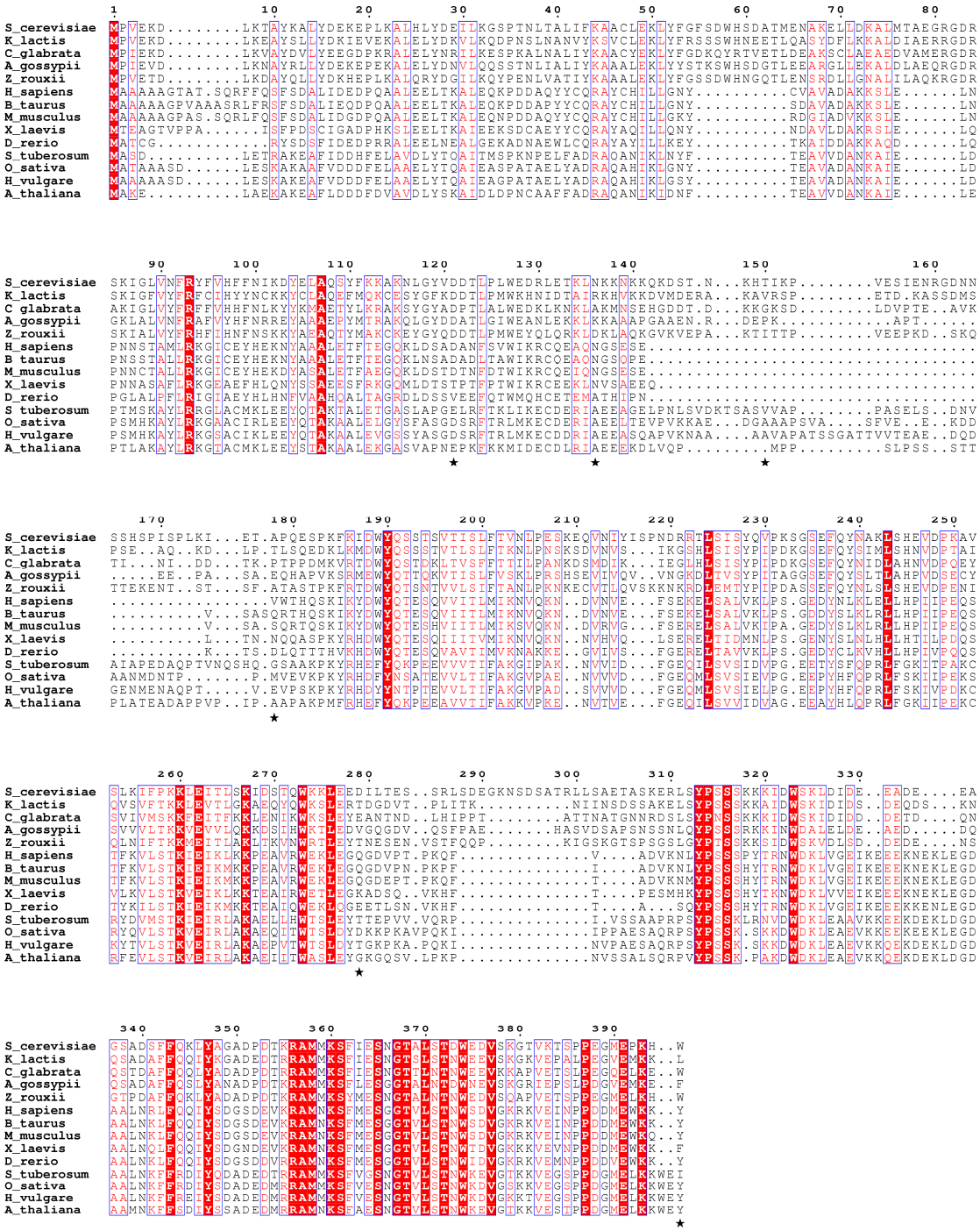
249. Zhang Y, Chan DC. 2007. Structural basis for recruitment of mitochondrial fission complexes by Fis1. Proc Natl Acad Sci USA. 104 (47): 18526-30.

250. Zhang Z, Roe SM, Diogon M, Kong E, El Alaoui H, Barford D. 2010. Molecular structure of the N-terminal domain of the APC/C subunit Cdc27 reveals a homo-dimeric tetratricopetide repeat architecture. J Mol Biol. 397 (5): 1316-1328.

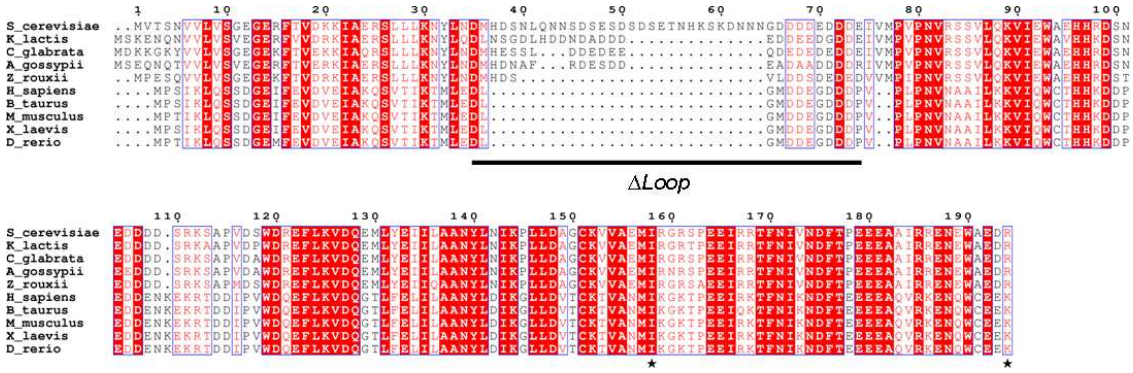
251. Zheng N, Schulman BA, Song L, Miller JJ, Jeffrey PD, Wang P, Chu C, Koepp DM, Elledge SJ, Pagano M, Conaway RC, Conaway JW, Harper JW, Pavletich NP. 2002. Structure of the Cul1-Rbx1-Skp1-F box Skp2 SCF ubiquitin ligase complex. Nature. 416 (6882): 703-709.

APPENDICES

APPENDIX A – STRUCTURE-BASED SEQUENCE ALIGNMENTS



Appendix A1: **Structure-Based Sequence Alignment of Sgt1 Homologues –** Residue cut-offs of specific truncations used in this study are marked by « . This structure-based alignment was generated using T-COFFEE EXPRESSO (Notredame *et al.*, 2000) and the figure made using ESPript (Gouet *et al.*, 1999).



Appendix A2: **Structure-Based Sequence Alignment of Skp1 Homologues -** Residue cut-offs of specific truncations used in this study are marked by « . This structure-based alignment was generated using T-COFFEE EXPRESSO (Notredame *et al.*, 2000) and the figure made using ESPript (Gouet *et al.*, 1999).

APPENDIX B – SUMMARY TABLE OF MASS SPECTROMETRY DATA

Species	Stoichiometry	Expected Mass (Da)	Observed Mass (Da)	Charge State	m/z
Sgt1 FL	1	44860	46867 ± 1.66	+12	3906.58
				+13	3606.15
				+14	3348.64
	2	89720	93734 ± 3.32	+17	5514.76
				+18	5208.44
				+19	4934.37
				+20	4687.70
				+21	4464.52
	3	134580	140601 ± 4.98	+22	6391.95
				+23	6114.09
				+24	5859.38
				+25	5625.04
				+26	5408.73
Sgt1ΔSGS (1-279)	1	33207	34119.05 ± 21.95	Data not available	
	2	66414	68242.34 ± 62.26		
	3	99621	102382.45 ± 50.76		
Sgt1 1-150	2	34924	34663.66 ± 6.46	+9	3852.52
				+10	3467.37
				+11	3152.24
				+12	2889.64
	3	96330	51995.49 ± 9.69	+11	4727.86
				+12	4333.96
				+13	4000.65
				+14	3714.96
				+15	3467.37
Skp1 FL	1	20970	22200.25 ± 4.04	+7	3172.43
				+8	2776.00
				+9	2467.67
				+10	2221.00
	2	41940	44400.5 ± 8.08	+11	4037.41
				+12	3701.04
				+13	3416.42
				+14	3172.46
Sgt1 FL:Skp1 FL	2:1	110690	115934 ± 7.36	+19	6102.80
				+20	5797.71
				+21	5521.68
				+22	5270.74
	1:1	65830	69067.25 ± 5.70	+15	4605.48
				+16	4317.70
				+17	4063.78
				+18	3838.07
Sgt1 D61R	1	47064	46997.57 ± 12.65	+13	3616.20
				+14	3358.00
				+15	3134.17
	2	94128	93995.14 ± 25.30	+19	4948.11
				+20	4700.76
				+21	4476.96
				+22	4273.51
				+23	4087.75
	3	141192	140992.70 ± 37.95	+25	5640.71
				+26	5423.80
				+27	5222.95
Sgt1 H59A	1	46957	46846.86 ± 8.91	+12	3904.91
				+13	3604.61
				+14	3347.20
				+15	3124.12
	2	93914	93693.72 ± 17.82	+19	4932.25
				+20	4685.69
				+21	4462.61
				+22	4259.81
				+23	4074.64
	3	140871	140540.58 ± 26.73	+25	5622.62
				+26	5406.41
				+27	5206.21

Appendix B1: **Summary of Mass Spectrometry Data**